



Bases génétiques et écologiques de la diversification adaptative chez *Escherichia coli*

Jessica Plucain

► To cite this version:

Jessica Plucain. Bases génétiques et écologiques de la diversification adaptative chez *Escherichia coli*. Sciences agricoles. Université de Grenoble, 2012. Français. NNT : 2012GRENV049 . tel-00947803

HAL Id: tel-00947803

<https://theses.hal.science/tel-00947803>

Submitted on 17 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Microbiologie - Virologie**

Arrêté ministériel : 7 août 2006

Présentée par **Jessica PLUCAIN**

Thèse dirigée par **Dominique SCHNEIDER** et co-encadrée par **Thomas HINDRE**

préparée au sein du **Laboratoire Adaptation et Pathogénie des
Microorganismes** (CNRS UMR5163)
dans l'**École Doctorale Chimie et Sciences du Vivant**

BASES GENETIQUES ET ECOLOGIQUES DE LA DIVERSIFICATION ADAPTATIVE CHEZ *ESCHERICHIA COLI*

Thèse soutenue publiquement le **11 Décembre 2012**,
devant le jury composé de :

Mr Thomas LENORMAND

DR, Montpellier, Rapporteur.

Mr Oliver TENAILLON

CR, Paris, Rapporteur.

Mr Guillaume BESLON

PU, Lyon, Examineur.

Mme Isabelle MEYNIAL-SALLES

MCU, Toulouse, Examineur.

Mme Irène TILL-BOTTRAUD

DR, Grenoble, Examineur.

Mr Dominique SCHNEIDER

PU, Grenoble, Directeur de thèse.

Mr Thomas HINDRE

MCU, Grenoble, Co-encadrant de thèse.



Ce manuscrit est dédié à Evelyne Coursange, notre technicienne au grand cœur et au large sourire, qui affronte les affres de la vie avec une volonté et un courage qui forcent l'admiration. Pussions-nous tous y trouver une source d'inspiration.

REMERCIEMENTS

Je remercie tout d'abord le Ministère de l'Enseignement Supérieur et de la Recherche et l'Université Joseph Fourier pour l'allocation de recherche et le poste d'enseignant contractuel qui m'ont été attribués pour la durée de ce doctorat. Ces travaux ont également été soutenus par deux financements ANR (programmes Génomique et Blanc).

Je remercie ensuite Guillaume Beslon, Isabelle Meynial-Salles, Thomas Lenormand, Oliver Tenaillon et Irène Till-Bottraud, membres de ce jury, pour consacrer de leur temps et de leurs compétences à l'évaluation de ce travail.

On demanda un jour à Cicéron : « Cicéron, qu'est-ce qu'un Grand Homme ? ». Ce à quoi il répondit : « Un Grand Homme est quelqu'un auprès de qui on se sent Grand ». Si pour ressortir scientifiquement grandi d'un doctorat il faut avoir été entouré d'Hommes scientifiquement Grands, de même, pour ressortir humainement grandi du même doctorat il faut avoir été entouré d'Hommes humainement Grands. Ces deux aspects étant indissociables à mes yeux, j'ai donc une myriade de remerciements à prodiguer aux Grands Hommes que j'ai eu la chance de rencontrer durant ce projet, et aux personnes qui ont participé de près ou de loin à cette aventure scientifique et humaine.

Si vous vous demandez s'il existe en la communauté scientifique un être brillantissime, enthousiaste, dynamique et qui pourrait de surcroît être attentionné envers ses congénères, désintéressé, doté d'un égo infime et profondément dévoué pour la communauté, ne cherchez plus, vous avez besoin de Dominique Schneider. Certes, il faut supporter également quelques sautes d'humeur les jours de pénurie de café, mais rien de dramatique. Dom je te remercie d'avoir cru en moi à un moment où peu de personne aurait parié sur moi et de m'avoir par la suite accordé une confiance sans limite. J'ai bénéficié, et je pense en mesurer pleinement la valeur, d'un encadrement exceptionnel majoré d'une liberté d'expression et d'agissement totale. Je te remercie également d'avoir ri de bon cœur à mes (légères) excentricités et mes (multiples) débordements humoristiques, supporté ma maniaquerie et résisté (jusqu'à maintenant) à l'envie de me crucifier sur la porte d'entrée du laboratoire.

Et si tout cela ne suffisait pas, j'ai eu la chance d'être co-encadrée par Thomas Hindré qui m'a enseigné avec une patience infinie (et il en faut) tout ce qu'il sait, au jour le jour. Thomas, je n'aurais pu mener à bien aucune des expériences dans lesquelles je me suis lancée si tu n'avais pris le temps de me les expliquer et montrer. Tes grandes capacités pédagogiques ont été pleinement appréciées, et m'ont profondément inspiré pour mes enseignements. Je te remercie également pour le temps que tu as consacré à mes écrits en tout genre. Et j'ai le secret espoir que, à ton contact, je me sois assagie (pourquoi souris-tu ?), ce qui n'est pas une mince affaire.

Joël Gaffé, écoute moi attentivement, je te dois approximativement un millier de remerciements. Merci pour ton infinie bonté, merci pour ta bienveillance, merci pour ton humilité, merci pour tes talents et tes connaissances en biochimie, merci pour ton aide précieuse avec les protéines ArcA, merci pour ton encadrement et tes conseils pour mes enseignements (je n'ai amoché aucun étudiant grâce à toi), merci pour ton humour et merci pour tes petits gâteaux, tes pastilles à la menthe qui arrachent, tes sachets de thé et tes aspirines.

A mon précieux compagnon de galère, Mickael Le Gac, je voudrais dire ma gratitude. J'ai adoré me sentir poussée à donner toujours plus, j'ai adoré les compétitions stupides de celui d'entre nous qui utilisera le plus d'erlens et de boîtes de pétri, j'ai adoré tes remarques assassines (mais tellement vraies) sur les contrôles manquant de mes expériences et j'ai adoré tes figures à multiples composantes incompréhensibles pour le commun des mortels. Merci pour l'inspiration et l'exemple.

Merci également à Sébastien Wielgoss, pour sa gentillesse, sa disponibilité, sa présence et pour les longues heures de discussion en tout genre qui m'ont obligé à progresser en anglais.

Un immense MERCI à Evelyne Coursange, gérante en chef d'à peu près tout, dont la présence maternelle nous manque terriblement dans le laboratoire. Me manque également son épaule pour pleurer les malheurs de la journée, ses pâtisseries à se damner, ses fameux coups de gueule du genre « Et le matériel, il se range tout seul ? », « Qui a fini le stock ? », « Je ne passe pas de commandes pour la veille ! ». Je suis infiniment reconnaissante d'avoir gagné ton amitié.

Au reste des membres de l'équipe, j'ai nommé Max Maurin, Antonia Suau-Pernet, Maxime Levert, Vivien Sutura, Pauline Manhes, Larissa Viraphong, Delphine Aymoz, Léa Beaume et Amandine Gaude, je vous sais gré de l'ambiance joviale du laboratoire, de mémorables fous rires, et de tous ces agréables moments passés en votre compagnie.

Une mention spéciale pour Colin Raeside, doctorant écossais barbu et très peu matinal, pour avoir supporté, dans la patience et la douceur, les taquineries plus ou moins douteuses que je lui ai infligées et pour avoir développé au cours du temps des répliques piquantes et hilarantes. Je pense également à tous les autres prodigieux « souffre-douleurs » qui l'ont précédé (Iyad, Mike, Omay, Julien...), cela n'aurait pas été pareil sans vous.

Un immense merci à Guillaume Baptiste et Jérôme Izard, pour avoir partagé la condition ingrate et fantastique à la fois de doctorant. Je les remercie également d'avoir veillé sur moi quand « celui-dont-on-ne-prononce-pas-le-nom » rodait dans les parages.

Ce doctorat n'aurait sans doute pas été possible sans la logistique remarquable de Christiane Vella, Bruno Thibaut et Séverine Sollier pour générer les quelques milliers de gélules, erlens et tubes utilisés.

Merci aux secrétaires pour leur formidable travail, et leur bonne humeur communicative.

Enfin, une pensée pour l'ensemble des membres de l'unité de recherche pour leur souhaiter un lumineux et harmonieux chemin pour la suite.

Je remercie également les membres extérieurs au laboratoire qui ont participé de près ou de loin au projet. J'ai eu la chance de côtoyer Olivier Tenaillon qui a toujours répondu à mes questions naïves sans jamais me déprécier et a été d'une grande aide dans les analyses des génomes, et Guillaume Beslon qui a porté un intérêt appréciable à mes recherches. Je les remercie tous deux pour les remarques constructives pour l'avancement de mes travaux, et pour ne pas m'avoir dénoncé (ou pas encore) pour trafic de signatures. Je remercie également Claudine Médigue et Stéphane Cruveiller pour leur travail exceptionnel effectué pour l'obtention des séquences de génomes, leurs compétences et leur gentillesse. Enfin je remercie Jacques Coves et Isabelle Petit-Hartlein pour m'avoir accueilli à l'IBS et m'avoir patiemment enseigné la purification de protéines.

L'ensemble de ce travail n'aurait pas été possible si, en 1988, Richard Lenski n'avait eu l'idée géniale et saugrenue de commencer cette expérience d'évolution au long terme. Je suis consciente de la chance exceptionnelle que j'ai eu de travailler sur ces lignées, et particulièrement sur la population Ara-2 qui recèle quelque chose de magique. Ce fut une joie de travailler sous l'aura d'un tel homme.

Je suis reconnaissante à Daniel Rozen pour l'immense travail qu'il a réalisé sur la population Ara-2, et pour m'avoir communiqué avec gentillesse les trucs et astuces issus de son expérience personnelle sur le sujet. Manier les S et les L est délicat, il le sait mieux que quiconque, et une sorte d'empathie ou de compassion mutuelle nous lie désormais. Merci pour tout Danny!

Un immense merci à Zachary Blount, pour avoir partagé avec moi ses précieuses connaissances en termes de spéciation et son expérience pour l'étude de populations exotiques. Merci Zack de m'avoir gratifié de ton amitié, de ton soutien et de tes blagues plus que douteuses sur les Français, et d'avoir été une excellente source d'inspiration pour mon travail.

J'ai également eu la chance d'enseigner dans diverses disciplines au cours de ce doctorat. Toute ma gratitude va donc aux enseignants qui m'ont aidé et intégré aux différentes équipes pédagogiques, aux étudiants qui ont été un formidable terrain d'expérimentation et grâce à qui j'ai beaucoup appris tant sur moi-même que sur la façon de gérer mes enseignements, ainsi qu'aux différents formateurs du CIES et à son directeur, Pierre Thibault, pour la passion vibrante qu'ils ont mis à transmettre leurs compétences pour l'enseignement supérieur. Quelle aventure ce fut, j'en ressors grandie.

Enfin je remercie mes camarades de paillasse préférés, j'ai nommé les S et les L. Que je vous aime, maudites bestioles, vous m'avez rendue folle, mais je ne peux me passer de vous maintenant...

Sur un plan plus personnel maintenant, merci à Arthur, Denver & co pour à peu près tout.

Merci à ma famille, et particulièrement à mes sœurs Mikhaëlla et Samaëlle pour leur soutien indéfectible au cours des années. Je vous aime infiniment...

Merci à Michelle Brousse et Macha Lioukovnikova qui ont été ma famille d'adoption sur Grenoble.

Merci à Raymond Cirio pour m'avoir insufflé l'amour de la biologie.

Merci à Mme Benimelli pour m'avoir donné les moyens de poursuivre mes études.

Merci à Mathieu Haddad pour m'avoir poussé à donner le meilleur de moi-même. Mathieu, ce doctorat n'existerait pas sans toi, j'ai adoré nos querelles puériles, nos concours en tout genre, ta mauvaise foi et ta maladresse caractérisée. Ton amitié m'a été tellement précieuse.

Merci aux êtres exceptionnels qui m'ont entouré de leurs ailes : Alexandre, Angela, Angélique, Anna, Aube, Audrey, Aurore, Bojidar, Deva, Elena, Estelle, Eviane, Fabrice, Florent, Francesco, Friedrich, Gamaliel, Heidi, Isabel, Isé, Jean, José, Katia, Lewis, Lioubomir, Liubov, Lucien, Paola, Philippe, Manou, Mathieu, Martin, Nicolas, Silen, Stella, Sylvia, Svétli...merci de votre amitié qui ne connaît ni les barrières du temps, ni celles de l'espace. Une pensée spéciale pour Aurélia Chaboudet, petite fée malicieuse des montagnes suisses, Jonathan de Nicolas pour avoir été un formidable compagnon d'exploration et de réflexion et Frédéric Lichère qui m'a permis de découvrir des qualités insoupçonnées, merci de ta confiance Frédéric.

Enfin, merci à OMA pour m'avoir inspiré et guidé vers la lumière, pour m'avoir permis de me trouver.

TABLE DES MATIERES

TABLE DES ILLUSTRATIONS	1
AVANT PROPOS	3
INTRODUCTION	7
 INTRODUCTION PARTIE I :	
LES RESEAUX DE REGULATION DE L'EXPRESSION DES GENES	8
<u>I. Les régulateurs globaux de la transcription</u>	10
1. L'ARN polymérase	10
2. Les protéines de type histone	13
<u>II. Les molécules signal</u>	17
1. (p)ppGpp	17
2. Le diguanylate cyclique	20
<u>III. Les systèmes à deux composants : exemple de ArcA/ArcB</u>	21
<u>IV. Les petits ARNs</u>	24
 INTRODUCTION PARTIE II :	
LES EVENEMENTS DE DIVERSIFICATION BACTERIENNE	26
<u>I. Les événements de diversification dans des environnements naturels</u>	28
1. Analyse de la diversité bactérienne dans les collections d'isolats naturels	29
2. Analyse de la diversité bactérienne chez les patients au court terme	30
a. Diversité lors d'infections extra-intestinales dues à <i>E. coli</i>	30
b. Les Small Colony Variants	32
3. Analyse de la diversité bactérienne chez les patients au long terme	34
<u>II. Les événements de diversification dans des environnements expérimentaux</u>	41
1. Environnements complexes générés par construction de niche	42
2. Environnements complexes générés par structuration spatiale	46
a. Ecologie	46
b. Génétique de la radiation adaptative	48
c. Evolution de la coopération	51
d. Evolution phénotypique	55
3. Environnements complexes générés par structuration nutritionnelle	57
 INTRODUCTION PARTIE III :	
L'EXPERIENCE D'EVOLUTION AU LONG TERME AVEC <i>ESCHERICHIA COLI</i>	61
<u>I. Le système expérimental</u>	62
<u>II. Parallélisme de l'évolution</u>	64
<u>III. Evolution non parallèle et innovations phénotypiques</u>	67
1. Capacité à utiliser le citrate	68
2. Emergence d'un polymorphisme au long terme	73

RESULTATS	79
RESULTATS PARTIE I : MAINTENANCE AU LONG TERME D'UN EVENEMENT DE DIVERSIFICATION BACTERIENNE	80
Article : Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with <i>Escherichia coli</i> .	
RESULTATS PARTIE II : BASES GENETIQUES DE L'EMERGENCE DE L'EVENEMENT DE DIVERSIFICATION BACTERIENNE	85
Manuscrit : Multi-step emergence of a long-term polymorphism in <i>Escherichia coli</i> .	
Résultats supplémentaires	140
<u>I. Qu'en est-il de la lignée L ?</u>	140
<u>II. Histoire évolutive de la lignée L</u>	142
RESULTATS PARTIE III : ANALYSES BIOCHIMIQUES ET MOLECULAIRES DE LA MUTATION DU GENE <i>arcA</i>	148
<u>I. Introduction</u>	149
<u>II. Matériel et Méthodes</u>	151
1. Milieux de cultures, souches et plasmides	151
2. Clonage des gènes <i>arcA</i> et surexpression des protéines ArcA	151
3. Purification des protéines ArcA	153
4. Mesure des interactions ADN-protéine	154
<u>III. Résultats</u>	156
1. Surexpression de ArcA	156
2. Purification des protéines ArcA	157
3. Interactions protéines-ADN	159
<u>IV. Conclusions</u>	161
DISCUSSION GENERALE	164
REFERENCES	180

TABLE DES ILLUSTRATIONS

Figure 1 : Régulation de l'expression, de la stabilité et de l'activité de RpoS	13
Figure 2 : Régulation de la superhélicité négative par Fis	16
Figure 3 : La réponse stringente	18
Figure 4 : Représentation schématique des réactions de transfert de phosphate dans le système à deux composants arcAB	24
Figure 5 : Illustration de la diversité phénotypique observée dans des isolats d' <i>E. coli</i> d'un patient infecté par un clone unique	31
Figure 6 : Phénotype de Small Colony Variants	33
Figure 7 : Points de prélèvement des isolats et durée de vie des patients	34
Figure 8 : Arbre phylogénétique retraçant les relations évolutives entre les différents clones	35
Figure 9 : Evolution parallèle dans les isolats CF173 et CF333	37
Figure 10 : Prélèvement de 112 isolats épidémiques de <i>B. dolosa</i> auprès de 14 patients atteints de mucoviscidose sur une période de 16 ans	38
Figure 11 : La phylogénie bactérienne révèle un réseau probable de transmission entre individus et entre organes	40
Figure 12 : Expériences de co-cultures démontrant que les trois clones CV101, CV103 et CV116 constituent un polymorphisme stable	43
Figure 13 : Structure du gène <i>acs</i> chez les clones d' <i>E. coli</i> ancestraux et évolués consommateurs d'acétate	45
Figure 14 : Diversité phénotypique, spécificité de niche et relations écologiques entre les cellules de <i>P. fluorescens</i> ayant évolué dans un environnement spatialement hétérogène	47
Figure 15 : Structure de l'opéron <i>wss</i>	49
Figure 16 : Détail des voies et loci affectés par des mutations dans les clones de phénotype WS	50
Figure 17 : Emergence de tricheurs à partir des clones WS au cours de l'évolution en environnement spatialement hétérogène	52
Figure 18 : Apparition de tricheurs et désintégration du polymère formé par les clones WS	53
Figure 19 : Evolution de la morphologie des colonies	56
Figure 20 : Profils de croissance de chaque type cellulaire sur milieu glucose-acétate	58
Figure 21 : Structure de l'opéron <i>aceBAK</i> et régulation négative par IciR	59
Figure 22 : Cycle de Krebs, shunt du glyoxylate, et voies métaboliques de consommation et/ou sécrétion d'acétate montrant des tendances d'expression différentes entre FS/ancêtre et SS/ancêtre	60

Figure 23 : Principe de l'expérience d'évolution au long terme	62
Figure 24 : Trajectoire de la valeur sélective au cours de 20 000 générations d'évolution expérimentale	64
Figure 25 : Expansion de la population Ara-3 au cours de l'évolution du phénotype Cit+	69
Figure 26 : La potentialisation permet d'augmenter le taux de mutation vers le phénotype Cit+	70
Figure 27 : Phylogénie de la population Ara-3	71
Figure 28 : Amplification en tandem dans les génomes des clones Cit+	72
Figure 29 : Emergence et maintien du polymorphisme	75
Figure 30 : Interactions négatives dépendantes de la fréquence entre les lignées S et L, mesurées par des expériences de compétition à court terme	75
Figure 31 : Croissance de clones S et L	76
Figure 32 : Croissance des S et des L dans des filtrats de cultures de clones S et L	77
Figure 33 : Expériences d'invasion réciproque entre d'une part des clones portant les mutations des gènes <i>spoT</i> , <i>arcA</i> et <i>gntR</i> et d'autre part l'ancêtre (a) ou un dérivé de l'ancêtre portant l'allèle évolué <i>spoT</i> (b)	142
Figure 34 : Expériences d'invasion réciproque contre 6,5S1	144
Figure 35 : Expériences d'invasion réciproque contre 6K3	144
Figure 36 : Croissance des clones 6K3 et 6,5S1 dans le surnageant de culture des clones 5000.5 (a), 6000.2 (b) et 6,5L4 (c)	147
Figure 37 : Surexpression des protéines 6His-ArcA-anc (a) et 6His-ArcA-S (b)	157
Figure 38 : Purification des protéines 6His-ArcA-anc (a et c) et 6His-ArcA-S surexprimées (b et d)	159
Figure 39 : Fixation des protéines 6His-ArcA-anc et 6His-ArcA-S sur les régions promotrices de <i>rpoS</i> (a) ou <i>fadE</i> (b)	161

AVANT PROPOS

La diversité biologique telle que nous l'observons aujourd'hui est le fruit de processus évolutifs complexes qui prennent racine dans un passé lointain. Au cours des dernières décennies, les recherches se sont multipliées pour tenter d'éclaircir les mécanismes d'émergence et de maintien de cette diversité, et ainsi améliorer notre compréhension de ces processus évolutifs et adaptatifs. Les études comparatives tout d'abord ont permis d'appréhender la complexité des ramifications de l'arbre de la vie et les relations de parenté qui unissent les êtres vivants. Puis ce sont les expériences d'évolution en laboratoire, au cours desquelles des organismes vivants s'adaptent à des conditions environnementales prédéfinies, qui ont permis d'étudier l'Evolution en action et les processus de diversification en direct. Bien qu'un certain nombre de ces études portent sur des organismes eucaryotes, les microorganismes, de par leur vitesse de multiplication, la taille de leur population, leur facilité de manipulation et les données physiologiques et moléculaires disponibles à leur sujet, sont devenus des modèles d'étude privilégiés de ces expériences d'évolution.

Les événements de diversification ainsi observés et étudiés présentent jusqu'alors un certain nombre de caractéristiques communes : 1. Ce sont des événements qui apparaissent de façon parallèle c'est-à-dire reproductible dans des environnements similaires à partir d'ancêtres communs. 2. Ces processus adaptatifs sont associés à une structuration spatiale de l'environnement ou à un mélange de ressources nutritionnelles, tel que défini par l'expérimentateur, ou suite à une modification de cet environnement par les organismes eux-mêmes. 3. Quand ils ont pu être mis en évidence, les

événements mutationnels impliqués dans cette diversification affectent des gènes (ou régions chromosomiques) uniques, et sont donc relativement facilement identifiables.

Les capacités évolutives et adaptatives des microorganismes impliquent un équilibre fin entre robustesse et évolvabilité et sont rendues possibles grâce à une plasticité du métabolisme ainsi que des réseaux de régulation géniques. Ainsi, les mutations permettant à ces organismes de s'adapter à de nouvelles conditions environnementales et de se diversifier, affectent fréquemment des gènes du métabolisme et/ou codant des régulateurs globaux, permettant par exemple l'accès à une source de carbone ou l'amélioration de son utilisation par une réorientation locale ou globale des fonctions cellulaires.

Ce projet se place au cœur des processus de diversification biologique d'une part, et de la flexibilité des réseaux métaboliques et géniques d'autre part. En effet, au cours de la plus longue expérience d'évolution en cours, où douze populations indépendantes sont propagées à partir d'un ancêtre commun d'*Escherichia coli* dans un environnement homogène et défini depuis plus de 55 000 générations, un événement unique de diversification a été détecté après 6500 générations dans une seule des douze populations : deux lignées, appelées S et L en référence à la taille de leurs colonies, ont émergé d'un ancêtre commun et continuent de co-exister depuis près de 50 000 générations. Certains des paramètres écologiques impliqués dans leur co-existence ont été identifiés et analysés de façon spécifique à 18 000 générations, et incluent une sélection négative dépendante de la fréquence et des taux de croissance et de déclin

différentiels. Cependant, les mécanismes moléculaires et génomiques de l'émergence et du maintien des deux lignées, ainsi que la dynamique des paramètres écologiques impliqués dans cet événement de diversification sont inconnus à ce jour.

Les travaux mis en œuvre au cours de ce doctorat se situent à deux niveaux pour l'étude de cet événement de diversification : d'une part, j'ai participé à l'étude du maintien du polymorphisme à long terme, notamment par l'analyse de l'évolution des interactions entre les deux lignées. D'autre part, mes travaux ont consisté à identifier à la fois les innovations phénotypiques amenant à cet événement de diversification, ainsi que les changements génétiques sous-jacents. En particulier, le chemin évolutif conduisant à l'établissement d'une des deux lignées a été mis en évidence. Nous avons ainsi pu montrer que trois mutations étaient nécessaires et suffisantes pour établir les phénotypes associés à l'une des lignées, dont les interactions négatives dépendantes de la fréquence avec l'autre lignée. Chacune de ces trois mutations affecte des régulateurs de la transcription des gènes. La spécificité des allèles sélectionnés, ainsi que l'effet moléculaire engendré par ces modifications ont été analysés. L'unicité de cet événement de diversification résulte d'une succession d'événements mutationnels précis, apparus et sélectionnés au cours de l'évolution.

INTRODUCTION

INTRODUCTION PARTIE I

LES RESEAUX DE REGULATION DE L'EXPRESSION DES GENES

Les microorganismes présentent de fantastiques capacités évolutives et adaptatives, qui leur ont permis de coloniser la quasi-totalité des niches écologiques présentes sur notre planète. Ces capacités évolutives résultent d'un équilibre subtil et modulable entre robustesse et évolvabilité, c'est-à-dire entre la capacité de résister à différentes perturbations (conditions environnementales, modifications génétiques) et celle d'évoluer et de s'y adapter. Ces notions de robustesse et d'évolvabilité peuvent s'appliquer à tous les niveaux cellulaires : biochimique, génétique et écologique (Lenski et al 2006). Cet équilibre résulte notamment de la plasticité des réseaux géniques des microorganismes, qui contrôlent de façon concertée d'une part les capacités métaboliques, et donc les capacités de croissance des cellules dans les différentes conditions de l'environnement, et d'autre part la réponse aux stress et aux différents changements environnementaux. Au cours d'expériences d'évolution en laboratoire, des mutations adaptatives sont fréquemment retrouvées dans les gènes codant les régulateurs clés de ces réseaux, qui constituent ainsi des cibles privilégiées de la sélection naturelle (Hindré et al 2012). De plus, une diversité naturelle importante a été observée au niveau de ces réseaux parmi les isolats environnementaux de différentes bactéries (Tenaillon et al 2010).

Une étude bibliographique exhaustive des réseaux géniques bactériens dépasse le cadre de ce travail. Nous restreindrons donc cette partie à la description de quelques exemples de régulateurs clés de ces réseaux géniques qui constituent les cibles de la sélection naturelle au cours d'expériences d'évolution en laboratoire et qui présentent une diversité dans les isolats naturels de certaines bactéries. Malgré cette forte restriction, cette description permettra de mettre en évidence que tous les niveaux de

régulation peuvent être affectés au cours des expériences d'évolution (voir Introduction Partie II section II). En effet, au cours de ces expériences, l'évolution des bactéries est associée à des modifications se produisant dans quatre des voies les plus courantes de régulation de l'expression des gènes. Des mutations ont ainsi été identifiées dans des gènes codant des régulateurs globaux de la transcription, des enzymes impliquées dans la synthèse de molécules signal, des systèmes à deux composants et des petits ARNs. Quelques exemples de chacun de ces quatre types de régulation globale sont donnés ci-dessous.

I. Les régulateurs globaux de la transcription

1. L'ARN polymérase

Le cœur de la machinerie transcriptionnelle, l'ARN polymérase, peut être la cible de la sélection naturelle. Ceci a notamment été démontré dans le cas de populations d'*E. coli* propagées dans du milieu minimum avec différentes sources de carbone (Herring et al 2006, Applebee et al 2008, Conrad et al 2010). Un certain nombre de mutations, bénéfiques dans ces conditions (Applebee et al 2008), ont été mises en évidence dans les gènes codant les sous-unités de l'ARN polymérase cœur. Elles altèrent les propriétés cinétiques de l'enzyme, en améliorant notamment la vitesse d'élongation de la transcription, et modifient la distribution des molécules d'ARN polymérase sur le chromosome (Conrad et al 2010). Ces modifications provoquent un changement global de l'expression des gènes, qui permet probablement aux bactéries une meilleure adaptation aux conditions de l'environnement, ici du milieu minimum de culture.

Dans d'autres expériences d'évolution où des populations d'*E. coli* ont été adaptées à des conditions de culture continue, des mutations affectent également un gène d'ARN polymérase, mais cette fois codant une sous-unité plus spécifique, celle correspondant au facteur sigma alternatif RpoS codé par le gène *rpoS* (Notley-McRobb et al 2002, Ferenci 2008, Wang et al 2010). Le facteur RpoS est exprimé dans des conditions de stress très variés, par exemple en phase stationnaire lors de carences nutritionnelles, en conditions de choc osmotique, de choc thermique, et de stress oxydatif. Le facteur sigma RpoS s'associe à l'ARN polymérase, en lieu et place du facteur sigma végétatif RpoD, ce qui la rend alors capable de transcrire des gènes nécessaires à la survie lors de ces différents stress. L'implication de RpoS dans une telle diversité de conditions de stress est réalisée grâce à une régulation très fine et complexe de son expression et ce, à différents niveaux (transcription, traduction, stabilité et activité protéique) en fonction des stimuli externes (Battesti et al 2011). La transcription de *rpoS* est soit activée par une petite molécule signal appelée (p)ppGpp (voir section II) et le système à deux composants BarA/UvrY, soit réprimée par le complexe CRP/AMPc et le système à deux composants ArcA/ArcB (voir section III) (Figure 1a). Par ailleurs, le transcrit majeur du gène *rpoS* forme une structure de type tige-boucle qui empêche la fixation des ribosomes. Lors de conditions de stress, différents petits ARNs spécifiques de chacun des stress (*dsrA*, *rprA*, *arcZ*) relâchent cette inhibition avec l'aide de la chaperone Hfq, et favorisent ainsi la traduction du transcrit *rpoS*. Un autre petit ARN, *oxyS*, inhibe la traduction de *rpoS*. D'autres protéines affectent également de façon positive (CsdA, CspC, CspE, HU, DksA) ou négative (RNase III, H-NS) la traduction de *rpoS* (Figure 1b).

Un troisième niveau de régulation affecte la stabilité de la protéine RpoS. En phase exponentielle de croissance, elle se fixe sur la protéine adaptateur RssB ce qui entraîne sa dégradation par la protéase ClpXP. En revanche, dans des conditions de stress, la cellule synthétise des protéines anti-adaptateurs (IraP, IraM, IraD), qui interfèrent avec RssB et conduisent à la stabilisation de RpoS. La protéine stable RpoS peut alors se fixer sur l'ARN polymérase cœur, ce qui la protège d'autant plus de la dégradation, et permettre la transcription des gènes nécessaires à la survie (Figure 1d). Enfin, un dernier niveau de régulation contrôle l'activité de la protéine RpoS. En effet, l'ARN 6S, les protéines Crl et Rsd, et la molécule signal (p)ppGpp favorisent la fixation du facteur sigma alternatif RpoS sur l'ARN polymérase cœur, en lieu et place du facteur sigma végétatif RpoD (Figure 1c). Chacun de ces régulateurs agit en fonction de l'un ou l'autre des stimuli ou stress auxquels la bactérie est confrontée. Certains, comme (p)ppGpp ou ArcA/ArcB, réagissent à de multiples stimuli.

Dans certaines expériences d'évolution en laboratoire où des populations d'*E. coli* sont propagées en cultures continues avec un apport régulier de nutriments, des mutations ont été retrouvées dans le gène *rpoS* (Notley-McRobb et al 2002). Ces mutations conduisent à une diminution de l'activité de RpoS permettant une meilleure croissance dans ces conditions où aucune carence nutritionnelle n'est présente. D'autre part, une diversité importante a été observée au niveau de la quantité de RpoS dans des isolats naturels d'*E. coli* issus de patients. Cette diversité est liée à des modifications de la régulation de RpoS (voir Introduction Partie II section I).

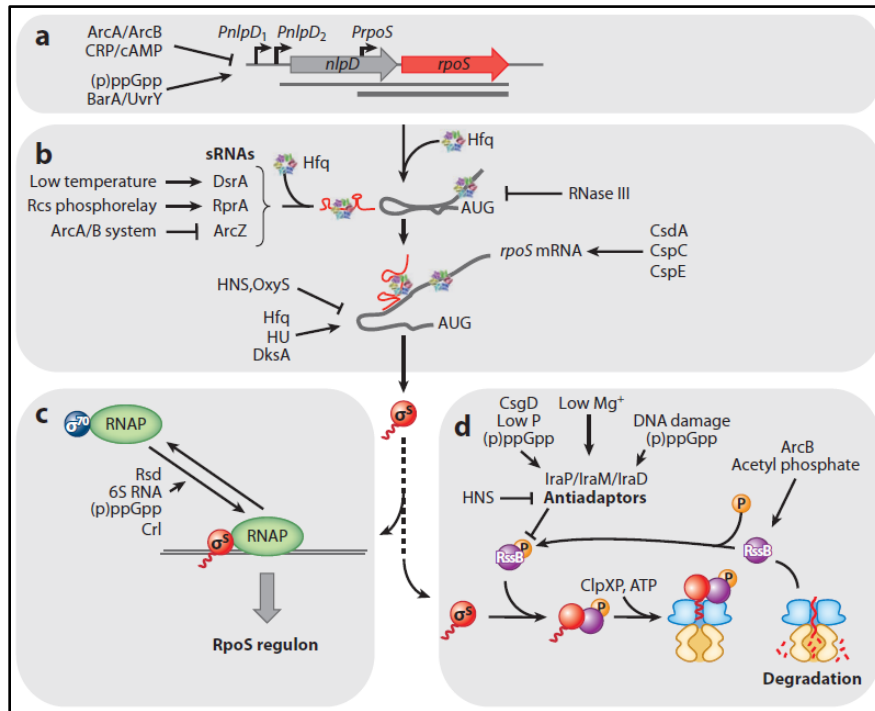


Figure 1 : Régulation de l'expression, de la stabilité et de l'activité de RpoS (d'après Battesti et al 2011). Les différents niveaux de régulation sont montrés dans les encadrés: **(a)** régulation transcriptionnelle, **(b)** régulation traductionnelle, **(c)** régulation de l'activité de RpoS, et **(d)** régulation protéolytique. Voir le texte pour les détails et les références relatifs à ces mécanismes de régulation.

2. Les protéines de type histone

Les protéines bactériennes de type histone comme Fis, HU, H-NS, IHF, Dps (Dorman et Deighan 2003), encore appelées NAP (Protéines Associées au Nucléoïde), interviennent dans la structure et l'état de superenroulement du chromosome (Luijsterburg et al 2006). Elles assurent la compaction et la condensation du chromosome en une structure appelée nucléoïde. En plus de cette activité d'organisation du génome, elles interviennent également dans son expression. En effet, de par leurs propriétés de liaison à l'ADN, elles constituent des facteurs de transcription

et contrôlent l'expression de centaines de gènes cibles, directement ou indirectement (Browning et al 2010). De par leur rôle conjoint dans le contrôle de la superhélicité de l'ADN et de la transcription globale du génome, les NAPs vont entraîner une dynamique importante du nucléoïde bactérien et de l'expression globale des gènes, en l'ajustant aux conditions de l'environnement. De plus, l'expression des NAPs est elle-même finement régulée, et elles sont exprimées différemment au cours du cycle cellulaire et en fonction de l'environnement. Les NAPs constituent donc des moyens de réponses aux fluctuations de l'environnement. C'est par exemple le cas de la protéine Fis qui est par ailleurs une des cibles de la sélection naturelle au cours de la plus longue expérience d'évolution en cours (Crozat et al 2005, 2010), où des cellules d'*E. coli* sont propagées en présence de glucose comme seule source de carbone (voir Introduction Partie II section II). De plus, la régulation des gènes par Fis est altérée selon les isolats d'*E. coli* (Crozat et al 2011).

Fis est une protéine fixant l'ADN dont l'expression est corrélée avec la phase de croissance. Elle est extrêmement abondante en phase exponentielle, puis sa quantité chute de manière drastique lors du passage en phase stationnaire (Ball et al 1992, Ninnemann et al 1992). La régulation du gène *fis* s'opère essentiellement au niveau transcriptionnel, son promoteur étant réprimé par un relâchement de la superhélicité de l'ADN (Schneider et al 2000) et par la molécule signal (p)ppGpp (Ninnemann et al 1992, voir section II.1), deux conditions qui sont remplies en phase stationnaire. Le promoteur de *fis* est également soumis à une auto-répression par Fis (Ball et al 1992, Finkel et Johnson 1992). Enfin, la régulation de l'expression de *fis* s'exerce également au niveau de sa traduction par une protéine à activité GTPase, appelée BipA qui interagit avec la

région 5' non traduite de l'ARNm de *fis* et les ribosomes (Owens et al 2004) et par la protéine DusB, par un mécanisme inconnu (Crozat et al 2010).

En plus de son rôle d'organisation du génome, Fis coordonne l'expression d'un grand nombre de gènes soit par contrôle direct de l'initiation de la transcription en se fixant sur des séquences spécifiques en amont des promoteurs, dont ceux permettant la transcription des gènes d'ARN ribosomiques et d'ARN de transfert (Bosch et al 1990, Ross et al 1990, Lazarus et Travers 1993), soit de façon indirecte en modulant la superhélicité du chromosome. Ces deux niveaux d'intervention sont par ailleurs intimement corrélés, car Fis est un régulateur direct des gènes de topoisomérases. Elle inhibe en effet la transcription des gènes *gyrA* et *gyrB* codant les sous-unités de la gyrase (Schneider et al 2000) et active le gène *topA* codant la topoisomérase I (Weinstein-Fischer et al 2000). Cet effet direct de Fis sur la transcription des gènes de topoisomérases entraîne une modulation de la dynamique de la superhélicité de l'ADN au cours de la phase de croissance (Figure 2), ce qui impacte de façon indirecte sur l'expression du génome dans son ensemble. L'expression de la protéine Fis étant très régulée en fonction de la phase de croissance, ces effets permettent donc de réguler le niveau d'organisation et d'expression du génome en fonction du cycle cellulaire.

II. Les molécules signal

1. (p)ppGpp

Lorsque les bactéries sont confrontées à des carences nutritionnelles, l'expression globale des gènes est modifiée de façon à inhiber leur croissance, économiser leur énergie et assurer leur survie. En particulier, la synthèse des protéines et des ARNs stables (ARNs ribosomiques et de transfert) diminue, tandis que la dégradation protéique et la synthèse d'acides aminés augmentent. Ce phénomène global est le résultat de l'intervention d'un réseau de régulation très complexe, appelé « réponse stringente », dont les effecteurs sont deux molécules signal : les nucléotides guanosine penta- et tétraphosphate, collectivement notés (p)ppGpp (Magnusson et al 2005). Lors de carences nutritionnelles, les cellules bactériennes synthétisent ces molécules grâce à l'activité de deux protéines, appelées RelA et SpoT. L'effecteur (p)ppGpp entraîne alors une reprogrammation de la transcription du génome qui a pour objectif l'arrêt de la croissance et la survie aux conditions de stress, avec notamment une inhibition de la transcription des ARNs stables, de la réplication de l'ADN, et une activation de la synthèse d'acides aminés et du facteur sigma alternatif RpoS (Figure 3). Quand les conditions redeviennent favorables, les molécules (p)ppGpp sont dégradées par SpoT qui possède donc les deux activités de synthèse et de dégradation de (p)ppGpp. La transcription du génome revient alors à une situation propice à la croissance bactérienne.

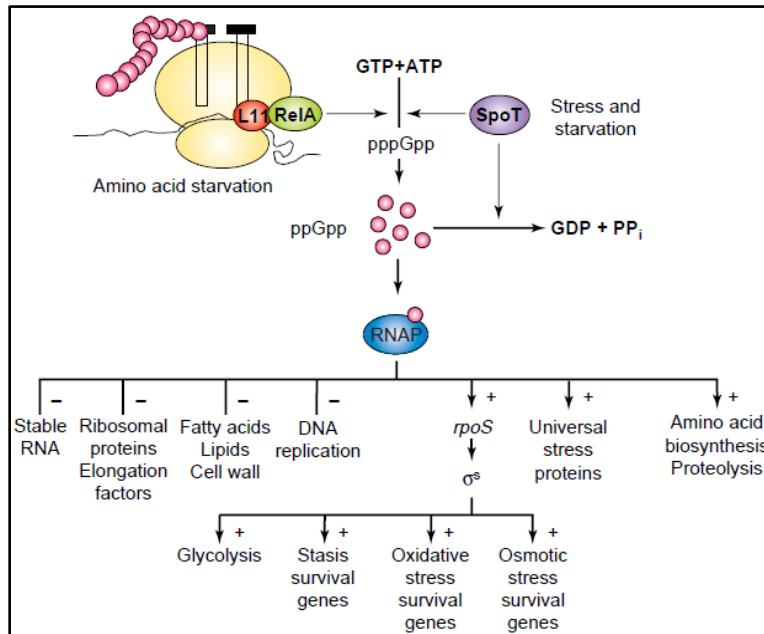


Figure 3 : La réponse stringente (d'après Magnusson et al 2005). (p)ppGpp est produit à partir de GTP et d'ATP par deux voies parallèles impliquant les enzymes RelA et SpoT, en réponse aux carences et aux stress, et est aussitôt converti en ppGpp. ppGpp se fixe à l'ARN polymérase (RNAP) et redirige la transcription de gènes dédiés à la croissance vers les gènes impliqués dans la résistance aux stress et la survie en phase stationnaire. La protéine SpoT est aussi responsable de l'hydrolyse du ppGpp.

La réponse stringente est donc basée sur la reprogrammation de la transcription globale du génome qui est réalisée via (p)ppGpp. La cible de (p)ppGpp est l'ARN polymérase, avec laquelle il interagit directement au niveau des sous-unités β et β' . Les effets de ces interactions sont de deux types. Tout d'abord, (p)ppGpp exerce un effet direct sur l'interaction entre l'ARN polymérase et les promoteurs régulés négativement. Lors de carences, l'accumulation de (p)ppGpp entraîne la déstabilisation des complexes d'initiation de la transcription au niveau de ces promoteurs, dont les promoteurs d'ARN stables (Barker et al 2001). Ces promoteurs étant ceux majoritairement transcrits en phase exponentielle, cette déstabilisation a pour effet indirect de libérer une grande quantité de molécules d'ARN polymérase. Celles-ci sont alors disponibles pour

transcrire des gènes dont les promoteurs ont une affinité moindre pour l'ARN polymérase, comme les gènes de biosynthèse et de transport d'acides aminés (Zhou et Jin 1998). Le deuxième effet de l'interaction de (p)ppGpp avec l'ARN polymérase est de diminuer l'affinité du facteur sigma végétatif RpoD pour l'ARN polymérase coeur, ce qui favorise la fixation des facteurs sigmas alternatifs dont la concentration est limitante (Jishage et al 2002). L'ARN polymérase est alors reprogrammée pour transcrire les gènes nécessaires à la survie aux stress.

La réponse stringente représente donc un des plus importants réseaux de régulation, qui est d'une part la cible de la sélection naturelle lors d'expériences d'évolution en laboratoire et qui présente d'autre part une forte diversité dans des isolats naturels bactériens. En effet, le gène *spoT* est une des cibles de la sélection naturelle au cours de différentes expériences d'évolution où des cellules d'*E. coli* sont propagées dans des conditions de cultures discontinues en milieu minimum (Cooper et al 2003, 2008, Ostrowski et al 2008, Woods et al 2011, voir Introduction Partie II section II). Les mutations du gène *spoT* se produisent en général de façon précoce au cours de l'évolution et confèrent un avantage sélectif qui est parmi les plus forts mesurés dans ce type d'expériences. Ces modifications de la réponse stringente pourraient favoriser les transitions entre phases du cycle cellulaire lors de ces cultures discontinues. Une diversité importante a également été mesurée dans les collections d'isolats naturels chez *E. coli* qui présentent une variabilité des quantités de (p)ppGpp et un polymorphisme au niveau des séquences nucléotidiques de *spoT* (Spira et al 2008, Ferenci et al 2011).

2. Le diguanylate cyclique

Le diguanylate cyclique (c-di-GMP) est un messager secondaire impliqué dans la régulation de nombreux processus physiologiques chez de nombreuses bactéries. Le c-di-GMP est synthétisé à partir de deux molécules de GTP par l'intermédiaire de diguanylate cyclases (DGC) et est hydrolysé par des phosphodiesterases (PDE) pour donner deux molécules de GMP. De nombreux génomes portent plusieurs gènes paralogues codant des DGCs et des PDEs dont les domaines catalytiques sont caractérisés par des motifs GGDEF et EAL ou HD-GYP, respectivement. Les gènes codant les enzymes DGCs sont des cibles pour l'adaptation et l'évolution de comportements altruistes chez les bactéries (Bantinaki et al 2007, McDonald et al 2009, voir Introduction Partie II section II).

Le c-di-GMP a été tout d'abord étudié dans le cadre de son rôle dans la régulation de la virulence chez des bactéries pathogènes telles que *Vibrio cholerae* ou *Pseudomonas aeruginosa*, chez lesquelles il active la production d'exopolysaccharides et de biofilm, et inhibe la motilité et la virulence (Tamayo et al 2007). Il est donc impliqué dans les phénomènes de transition entre différents modes de vie, sessiles et mobiles. Plus récemment, il a été montré que le rôle de messager secondaire du c-di-GMP était encore plus général. En effet, il est notamment impliqué dans la modulation de la croissance des bactéries sur des surfaces biotiques ou abiotiques, en régulant l'expression des composants nécessaires à l'adhésion cellulaire et en préparant les bactéries aux interactions cellule-cellule et cellule-surface (Jenal 2004). Le c-di-GMP est donc également impliqué dans le contrôle de l'expression de facteurs de virulence.

Les mécanismes par lesquels le c-di-GMP contrôle les différents processus cellulaires dans lesquels il est impliqué sont multiples et pas toujours clairement identifiés (Ryan et al 2012). L'exemple le mieux décrit implique un domaine de fixation du c-di-GMP, appelé PilZ, identifié dans de nombreuses protéines (Amikam et Galperin 2006). La fixation du c-di-GMP à ce domaine induit un changement conformationnel (Shin et al 2011) qui modifie l'activité et/ou la fonction des protéines possédant ce domaine PilZ. Dans d'autres cas, ce sont des protéines possédant des motifs GGDEF et/ou EAL ayant perdu leur activité catalytique qui assurent la transduction du signal lié au c-di-GMP (Newell et al 2009). Enfin, le c-di-GMP peut aussi agir comme régulateur transcriptionnel en se fixant soit à des facteurs de transcription dont l'activité est alors modifiée (Chin et al 2010), soit directement à la région non traduite d'ARNm (Sudarsan et al 2008).

La concentration de c-di-GMP est soumise à un contrôle drastique. L'expression des enzymes contrôlant son métabolisme (DGCs et PDEs) est régulée à une multitude de niveaux : la transcription des gènes codant ces enzymes, la stabilité des DGCs et PDEs, et leur activité qui est modulée par des phénomènes d'allostérie et de modifications post-traductionnelles comme des phosphorylations (Tamayo 2007). De tels niveaux fins et multiples sont requis au vu de l'importance et de la diversité des processus régulés par le c-di-GMP.

III. Les systèmes à deux composants : exemple de ArcA/ArcB

Les systèmes à deux composants (SDC) sont des régulateurs qui permettent d'intégrer directement des signaux de l'environnement, puis de les transmettre pour moduler

l'expression des gènes. Ils permettent ainsi aux cellules de répondre de façon appropriée aux stimuli de l'environnement. Généralement, les SDCs sont constitués de deux protéines (bien que des variantes plus complexes existent) : une kinase senseur et un régulateur de réponse (Mikkelsen et al 2011). La kinase senseur est en général une protéine membranaire (mais peut être cytoplasmique), qui est capable de détecter des signaux environnementaux (nutriments, substances toxiques, changements de pH ou de température, ...). Suite à cette détection, la kinase senseur s'auto-phosphoryle au niveau d'un résidu histidine conservé. Le groupement phosphate est ensuite transféré au régulateur de réponse, au niveau d'un résidu aspartate. Cette phosphorylation induit alors l'activité de ce régulateur de réponse qui agit comme un facteur de transcription et peut activer ou réprimer la transcription de ses gènes cibles. La cellule est alors capable de réagir aux stimuli de son environnement en ajustant l'expression de ses gènes. Les SDCs interviennent dans de nombreux processus cellulaires, dont les réponses aux stress, la fixation d'azote, la virulence, la compétence bactérienne, le chimiotactisme, la formation de biofilms, ...

Le seul système à deux composants connu à ce jour pour avoir été la cible de la sélection naturelle au cours d'expériences d'évolution est le système ArcA/ArcB (Le Gac et al 2010, ce travail). Le SDC ArcA/ArcB est un système de transduction du signal qui joue un rôle clé dans la régulation transcriptionnelle du métabolisme bactérien. Il permet notamment aux bactéries anaérobies facultatives de détecter différentes conditions d'oxygénation et d'y répondre en ajustant l'expression de leurs gènes (Lynch et Lin 1996). Le système ArcA/ArcB est composé de la kinase senseur ArcB, associée à la membrane, et du régulateur de réponse ArcA qui est un facteur de transcription. Le

signal reconnu par le système ArcA/ArcB correspond à l'état redox de la cellule par l'intermédiaire du pool de la forme oxydée des transporteurs d'électrons de type quinones de la chaîne respiratoire (Georgellis et al 2001). Dans des conditions d'anoxie, c'est-à-dire de diminution de la concentration en oxygène, ArcB s'auto-phosphoryle puis trans-phosphoryle ArcA (ArcA-P), qui régule alors la transcription de plus de 100 gènes cibles (Salmon et al 2005, Malpica et al 2006) (Figure 4). La protéine ArcA est notamment responsable dans ces conditions de la répression des gènes codant les enzymes majeures du cycle de Krebs et du shunt du glyoxylate (Wolfe 2005), ainsi que du gène *rpoS* (Hengge 2008). Un certain nombre de gènes intervenant dans le métabolisme sont également activés. Dans des conditions aérobies, non seulement l'activité de phosphorylation de ArcB est inhibée, mais ArcB agit cette fois comme une phosphatase en déphosphorylant ArcA-P (Pena-Sandoval et al 2005), ce qui relève la répression des gènes des enzymes du cycle de Krebs. En aérobie, le signal déclenchant l'activité phosphatase de ArcB est constitué par les transporteurs d'électrons de type quinones (Georgellis et al 2001). Le SDC ArcA/ArcB permet ainsi de faire le lien entre l'état redox de la cellule et le métabolisme central. De plus, ce SDC étant un régulateur de la transcription du gène *rpoS*, cela en fait un des régulateurs les plus globaux de la transcription. Le système ArcA/ArcB procure donc une plate-forme permettant aux cellules bactériennes d'adapter la transcription globale des gènes et l'activité des voies métaboliques en fonction de l'état redox et de la disponibilité en oxygène.

Le gène *arcA* est la cible de la sélection naturelle dans des populations d'*E. coli* propagées dans un milieu contenant deux sources de carbone différentes et est

impliqué dans la diversification bactérienne (Le Gac et al 2010, voir Introduction Partie II section II). Au cours de ce travail, nous avons également démontré le rôle crucial joué par ce régulateur dans l'émergence d'un événement de diversification bactérienne.

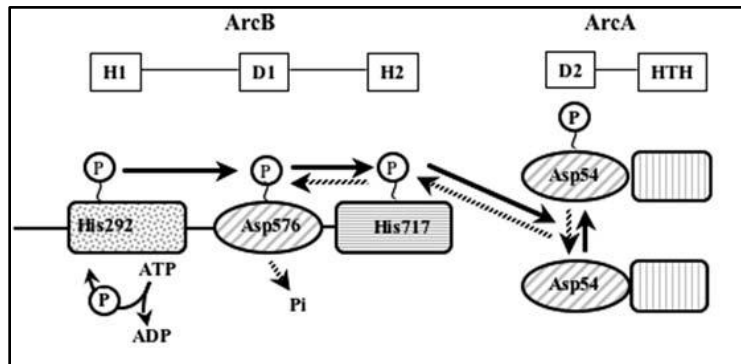


Figure 4 : Représentation schématique des réactions de transfert de phosphate dans le système à deux composants ArcA/ArcB (d'après Malpica et al 2006). Les flèches pleines indiquent la voie de transmission du signal amenant à la phosphorylation de ArcA. Les flèches hachurées indiquent la voie de coupure du signal amenant à la déphosphorylation de ArcA. Les acides aminés impliqués dans les réactions de transfert de phosphate sont en gras.

IV. Les petits ARNs

L'identification de petits ARNs régulateurs chez les bactéries a littéralement explosé depuis une décennie. De nombreuses bactéries produisent de petits transcrits, appelés petits ARN (sARN) et typiquement constitués de 50 à 300 nucléotides, souvent codés dans des régions intergéniques (Storz et al 2011). Ces sARNs peuvent réguler l'expression de leurs gènes cibles par plusieurs mécanismes. Un premier type implique la présence chez les sARNs d'une séquence nucléotidique complémentaire de leurs ARNm cibles, ce qui permet l'appariement entre sARNs et ARNm cibles. La séquence complémentaire sur les ARNm est souvent localisée au niveau du site de fixation des

ribosomes, la fixation du sARN bloquant ainsi la traduction (Gottesman 2004). Dans certains cas, le site d'interactions se situe à un autre endroit de l'ARNm et interfère avec sa stabilité. Chez de nombreuses bactéries à Gram-négatif, l'appariement de cette famille de sARNs requière la protéine chaperone Hfq. D'autres sARNs influent sur l'expression des gènes en se liant directement à des protéines, ce qui modifie leur activité ou les séquestre (Babitzke et Romeo 2007, Storz et al 2011).

Les petits ARNs et la machinerie de régulation qui les accompagne ont été les cibles de la sélection naturelle lors d'expériences d'évolution. En effet, des mutations bénéfiques ont été mises en évidence dans *hfq* lorsque des populations d'*E. coli* ont été propagées dans des conditions de cultures continues en chémostat (Maharjan et al 2010, Wang et al 2010, voir Introduction Partie II section II). Par ailleurs, une mutation a été identifiée dans un gène codant un sARN dans une expérience d'évolution avec *Myxococcus xanthus* (Yu et al 2010, voir Introduction Partie II section II). L'analyse de cette mutation a permis d'identifier la fonction de ce sARN, qui est impliqué dans le contrôle du développement de *M. xanthus* (Yu et al 2010).

Toute cette partie montre bien la sophistication des réseaux de régulation de l'expression des gènes, ainsi que la diversité des types de régulation et de leur niveau d'action. Tous ces niveaux et types de régulations peuvent être les cibles de la sélection naturelle et peuvent donc être encore affinés pour conférer aux bactéries une meilleure compétitivité dans leur environnement.

INTRODUCTION PARTIE II

LES EVENEMENTS DE DIVERSIFICATION BACTERIENNE

Une des caractéristiques majeures du monde vivant est son extrême diversité et complexité. Il résulte de multiples événements de spéciation qui se sont produits dans un passé lointain et résultent souvent de processus uniques, auxquels nous n'avons malheureusement pas accès. Ces processus ne nous offrent, pour leur étude, que leurs produits, à savoir les espèces contemporaines que l'Homme classe selon des critères qui peuvent être remis en question au fur et à mesure des recherches scientifiques. Au cours du temps ont co-existé différents concepts pour définir ce qu'est une espèce, la plupart étant issus et /ou appliqués aux organismes eucaryotes multicellulaires. Chez les organismes bactériens, la notion même d'espèce est difficile à définir, et les concepts qui en découlent restent un sujet de discussion complètement ouvert.

Le concept d'espèce biologique tel qu'il est défini chez les organismes sexués est dû à Ernst Mayr qui la définit comme « un groupe de populations capables de se reproduire entre elles pour donner une descendance fertile, et qui sont isolés d'autres groupes au niveau reproductif » (Mayr 1963). Bien évidemment, cette définition devient difficile à appliquer aux bactéries, du fait de l'absence de reproduction sexuée et de lien direct entre recombinaison et reproduction sexuée. En conséquence, d'autres concepts ont fleuri pour définir une espèce microbienne, certains allant même jusqu'à suggérer l'absence d'espèces bactériennes.

Un des concepts se rapprochant de la définition donnée par Ernst Mayr provient de Fred Cohan qui a proposé de transférer la notion d'espèce vers celle d'écotype (Cohan 2002, Cohan et Koeppel 2008), c'est-à-dire un groupe de bactéries occupant la même niche écologique, au sens large. Une espèce est alors définie comme l'équivalent d'un écotype, qui forme un groupe monophylétique. La spéciation pourrait alors se produire

lorsque des mutations permettant d'accéder à une nouvelle niche écologique sont fixées chez certains individus, leur donnant ainsi l'opportunité de se déplacer de la niche écologique ancestrale vers une nouvelle. Puisque ce nouveau groupe occupe une nouvelle niche, celui-ci et le groupe ancestral peuvent maintenant avoir un avenir évolutif indépendant, c'est-à-dire subir des événements de sélection périodique indépendants. Ces deux écotypes peuvent ainsi diverger indépendamment et fixer par exemple des mutations adaptatives différentes. Ce concept fait ainsi le parallèle entre l'isolement reproducteur défini par Ernst Mayr et l'isolement évolutif. Ainsi, il sera possible de détecter de tels événements de diversification, qui eux-mêmes pourraient ensuite aboutir à des événements de spéciation, sur la base des génotypes et des phénotypes spécifiques à chacun des écotypes.

Dans cette partie, nous étudierons tout d'abord les événements de diversification bactérienne qui ont été observés et étudiés dans des environnements naturels, avant de passer à ceux qui ont été détectés dans des environnements expérimentaux contrôlés en laboratoire.

I. Les événements de diversification dans des environnements naturels

Les événements de diversification en conditions naturelles sont difficiles à appréhender, car il est impossible de les anticiper pour pouvoir les observer et les analyser. La recherche s'est donc focalisée dans un premier temps sur des études comparatives en analysant les isolats naturels présents dans des collections de souches. Plus récemment, des événements de diversification ont été observés directement chez des

patients au cours d'infections bactériennes par *E. coli*, c'est-à-dire sur des durées de temps très courtes. D'autre part, la capacité des bactéries à engendrer de la diversité, tant au niveau génétique que phénotypique, au cours d'une infection a maintenant été identifiée chez de nombreux pathogènes, sous la forme de variants à petites colonies (Small Colony Variants – SCVs), et décortiquée au niveau génétique pour certains d'entre eux. Un troisième type d'études a émergé plus récemment encore, suite au suivi de patients au sein d'unités hospitalières pendant plusieurs dizaines d'années au cours desquelles les bactéries ont été isolées de ces patients, ce qui a offert l'opportunité de suivre l'évolution et l'adaptation de certains pathogènes directement dans l'environnement naturel de l'hôte sur des durées de temps beaucoup plus longues.

1. Analyse de la diversité bactérienne dans les collections d'isolats naturels

Des analyses comparatives de différentes souches présentes dans des collections de bactéries pathogènes et commensales d'*E. coli* ont été réalisées. Ces études ont permis de montrer une diversité importante au sein de cette espèce, avec l'identification de quatre groupes phylogénétiques majoritaires, A, B1, B2, et D, et de deux groupes accessoires C et E (Tenaillon et al 2010). Cette classification est basée sur plusieurs techniques : MLEE (multilocus enzyme electrophoresis), sérotypage, MLST (multilocus sequence typing), RFLP (restriction fragment length polymorphism) et séquençage de génome (Tenaillon et al 2010). Des études plus récentes ont révélé que cette diversité était également visible au niveau de l'expression différentielle de régulateurs globaux (voir Chapitre I) tels que RpoS (King et al 2004, Bhagwat et al 2006, Ferenci et al 2011)

et de la quantité de métabolites régulateurs comme ppGpp (Ferenci et al 2011) au sein d'isolats naturels d'*E. coli*. Ces variations conduiraient à un ajustement fin du niveau d'expression de leurs gènes cibles, permettant ainsi aux bactéries de trouver le compromis optimal entre d'une part les conditions de l'environnement et d'autre part leurs capacités métaboliques et leur tolérance aux stress rencontrés (King et al 2004, Ferenci 2005). Cette diversité au niveau des réseaux de régulation a également été retrouvée chez les bactéries isolées directement de patients (voir paragraphe 2 ci-dessous). Une diversité génétique importante a également été observée au niveau du gène *rpoS* dans une collection d'isolats cliniques de *Salmonella* (Robbe-Saul et al 2003).

2. Analyse de la diversité bactérienne chez les patients au court terme

a. Diversité lors d'infections extra-intestinales dues à *E. coli*

La diversité génomique et phénotypique de 226 isolats d'*E. coli*, isolés de 19 patients lors d'infections extra-intestinales, a été analysée (Levert et al 2010). Une variabilité génomique a été observée chez 11 de ces patients, cette diversité pouvant résulter d'infections polyclonales ou à la suite de micro-hétérogénéités apparaissant suite à des infections monoclonales. Ces micro-hétérogénéités lors d'infections monoclonales se caractérisent par une diversité phénotypique remarquable, affectant des phénotypes comme la résistance aux antibiotiques et aux stress, la perméabilité membranaire, le taux de croissance (Figure 5). Ce niveau de diversité n'est pas retrouvé dans le cas d'isolats commensaux d'*E. coli*, suggérant ainsi que des processus de diversification

spécifiques peuvent survenir lors du processus infectieux. De plus, cette diversité phénotypique est corrélée avec une variabilité des quantités de RpoS dans ces souches. Le séquençage du génome de clones d'*E. coli* isolés de ces patients n'a cependant révélé aucune mutation dans le gène *rpoS*, ce qui suggère que des gènes régulateurs de *rpoS* ont été modifiés au cours de l'infection (voir Introduction Partie I section I.1), aboutissant ainsi à des quantités variables de RpoS.

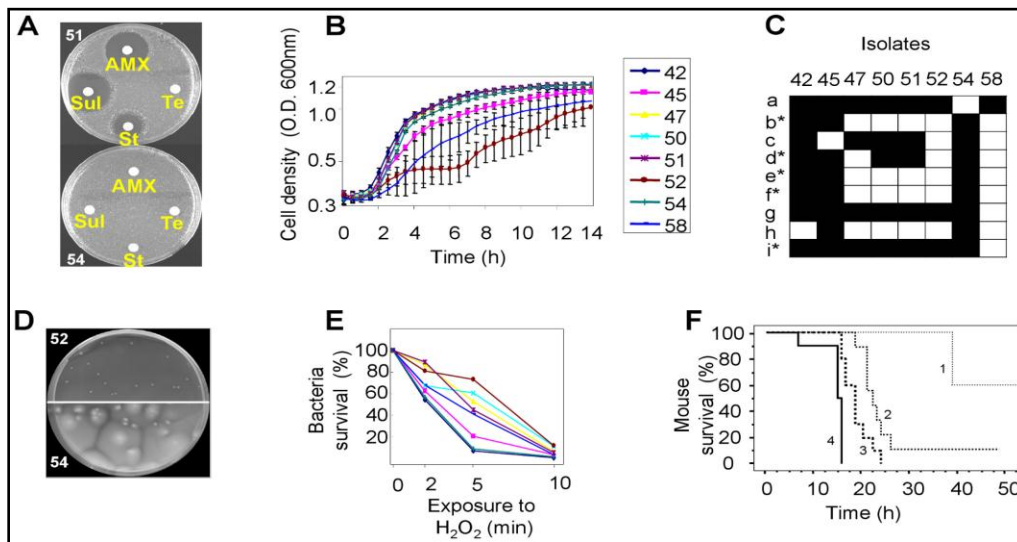


Figure 5 : Illustration de la diversité phénotypique observée dans des isolats d'*E. coli* [isolats issus du sang (n°42–50) et d'abcès du foie (n°51-58)] d'un patient infecté par un clone unique (d'après Levert et al 2010). (A) Variabilité des phénotypes de sensibilité aux antibiotiques. AMX: amoxicilline, Sul: sulfonamide, St: streptomycine, Te: tétracycline. (B) Courbes de croissance en milieu LB à 37°C pour huit isolats. (C) Variabilité de croissance sur 95 substrats (plaques Biolog GN2). Seuls les substrats différentiellement utilisés par les isolats sont indiqués : a: dextrines, b: acide bromosuccinique, c: glucuronamide, d: L-alanine, e: L-asparagine, f: L-acide aspartique, g: acide glycyl-L-aspartique, h: L-serine, i: D,L- α -Glycérol Phosphate. Les substrats avec une étoile correspondent aux substrats pour lesquels le métabolisme est stimulé par une perturbation de *rpoS*. Les carrés noirs et blancs indiquent l'utilisation ou l'absence d'utilisation du substrat, respectivement. (D) Variabilité de la motilité. Isolats non-motile (52) et motile (54). (E) Variabilité de la sensibilité à H₂O₂ mesurée par le pourcentage de survie à différents temps. Le code couleur des isolats est le même qu'en (B). (F) Variabilité au niveau de la survie de souris inoculées par différents isolats.

Ainsi, une variabilité importante des réseaux de régulation a été mise en évidence au cours du processus infectieux, comme cela avait été démontré lors d'analyses comparatives dans les collections d'isolats naturels.

b. Les Small Colony Variants

Depuis leur première mise en évidence en 1910, des variants formant de petites colonies (Small Colony Variants - SCVs) ont été détectés de façon répétée dans la descendance de nombreuses bactéries pathogènes, telles que *Salmonella typhi*, *E. coli*, *Shigella*, *Vibrio*, *Bacillus*, *Corynebacterium*, *Serratia marcescens*, *Neisseria gonorrhoeae*, *P. aeruginosa*, *Burkholderia cepacia*, *Stenotrophomonas maltophilia*, *Streptococcus pneumoniae* et *Staphylococcus aureus* (Melter et Radojevič 2010). Bien que les caractéristiques de tels variants soient multiples, il est tout de même possible de dégager des points communs aux SCVs. En effet, ils constituent généralement une sous-population qui émerge après l'infection d'une cellule hôte par une bactérie pathogène donnée, et qui présente des caractéristiques phénotypiques et pathogéniques distinctes par rapport à la souche de départ. Les phénotypes associés aux SCVs incluent une croissance lente et une déficience dans les systèmes de transport d'électrons (Proctor et al 1998), ce qui leur confère une petite taille de colonies après étalements sur boîtes de Pétri (Figure 6).



Figure 6 : Phénotype de Small Colony Variants (d'après Geipel 2009). Croissance sur gélose au sang de la bactérie *Staphylococcus aureus*. Des colonies de type sauvage (WT) et Small Colony Variant (SCV) sont visibles.

Les SCVs présentent fréquemment une résistance accrue aux antibiotiques en général, mais plus particulièrement aux aminoglycosides, une résistance au sérum de l'hôte (Häussler et al 2003), une hyper-adhérence aux cellules hôtes (Malone et al 2010), et une meilleure persistance intracellulaire (Proctor et al 1995). L'ensemble de ces phénotypes leur confère un avantage sélectif important par rapport aux cellules sauvages et un rôle important au cours des infections chroniques, notamment de par leur persistance. Les bases génétiques de l'émergence de ces variants sont diverses, et chaque nouvelle étude apporte son lot de gènes mutés, en fonction de l'espèce bactérienne, du lieu d'infection, du type de patient, etc. Il est tout de même important de noter que les SCVs présentent fréquemment un phénotype mutateur et sont systématiquement affectés dans les voies de biosynthèse des thiamines, thymidines, hémines et/ou ménadiones (Besier et al 2007, Chatterjee et al 2008, Lannergard et al 2008), soit par des mutations dans les gènes codant des enzymes clés, soit par dérégulation de ces voies. Ceci altère alors les systèmes de transport d'électrons et/ou le fonctionnement du cycle de Krebs, et donc la production d'ATP dans la cellule

(Chatterjee et al 2007). Des travaux récents ont également montré le rôle de la réponse stringente (voir Introduction Partie I Section II.1) dans l'établissement des phénotypes associés aux SCVs (Gao et al 2010). Ainsi, la réorganisation du métabolisme central et des réseaux de régulation semble représenter une des clés de l'émergence de ces variants dans le règne bactérien.

3. Analyse de la diversité bactérienne chez les patients au long terme

Deux études récentes ont eu pour objectif le suivi au long-terme de patients atteints de mucoviscidose et infectés par *P. aeruginosa* dans un cas (Yang et al 2011) et *Burkholderia dolosa* dans l'autre (Lieberman et al 2011). La diversité bactérienne générée au cours du temps a été analysée.

Un ou plusieurs prélèvements de *P. aeruginosa* ont été effectués entre 1973 et 2008 sur les expectorations de six patients atteints de mucoviscidose au Danemark (Yang et al 2011, Figure 7).

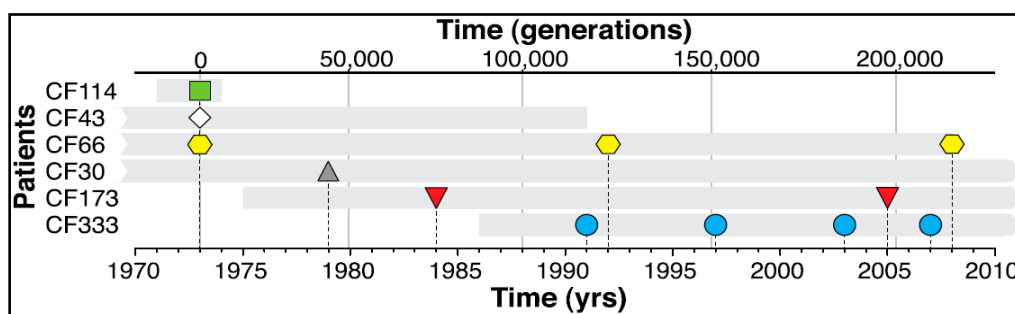


Figure 7 : Points de prélèvement des isolats et durée de vie des patients (d'après Yang et al 2011). Des isolats de *P. aeruginosa* ont été collectés de six patients différents souffrant de mucoviscidose sur une période de 35 ans. Les différents isolats bactériens sont représentés par des symboles de couleur (une couleur par patient), et les bandes grises représentent la durée de vie des patients.

Le génome des souches isolées a été séquencé et l'arbre phylogénétique qui en découle a permis de déterminer leurs relations de parenté. Au sein des prélèvements effectués, ceux datant d'avant 1979 semblent plus hétérogènes que ceux ultérieurs (Figures 7 et 8).

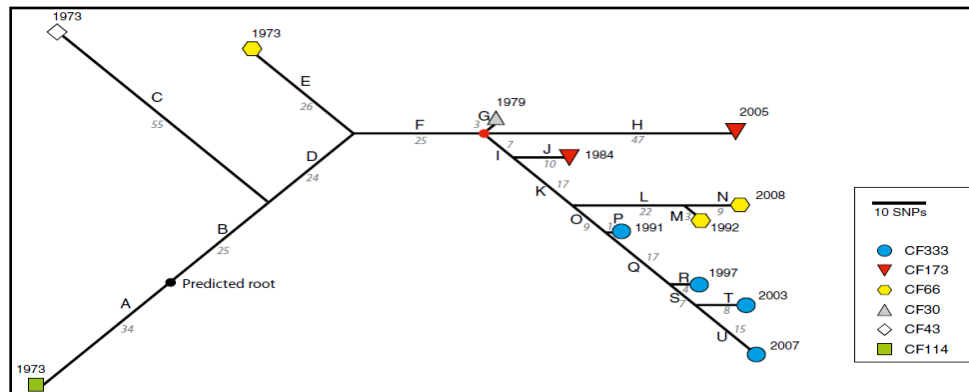


Figure 8 : Arbre phylogénétique retraçant les relations évolutives entre les différents clones (d'après Yang et al 2011). L'arbre est basé sur l'accumulation de mutations ponctuelles (SNP) identifiées par séquençage des génomes. La longueur des branches est proportionnelle aux nombres de SNPs (indiqués en gris italique) entre les différents isolats. Le cercle noir qui sépare les branches "A" et "B" représente la racine prédite de l'arbre, c'est-à-dire l'ancêtre commun immédiat des différents isolats.

Ces analyses génomiques montrent également que le taux d'évolution génomique augmente de façon constante au cours du temps, contrairement au taux d'évolution phénotypique qui augmente fortement dans les premières années suivant l'infection (avant 1979), signe de l'adaptation des bactéries à leur hôte, puis se stabilise (Yang et al 2011). Ces résultats sont en adéquation avec ceux observés au cours de la plus longue expérience d'évolution en laboratoire (Barrick et al 2009). Il est à noter que c'est au cours de la phase d'adaptation des bactéries (isolées avant 1979) qu'ont été identifiées le plus de mutations affectant des gènes codant des régulateurs globaux de la transcription, comme *mucA* (codant un anti facteur sigma), *lasR* (codant un régulateur

transcriptionnel impliqué dans le phénomène de quorum-sensing), et *rpoN* (codant le facteur sigma-54 de l'ARN polymérase), tous trois impliqués dans la régulation de la virulence chez *P. aeruginosa* (Cezairliyan et Sauer 2009, Sappington et al 2011, Damron et al 2012). Confirmant ces modifications génétiques, plus de la moitié des changements d'expression de gènes détectés dans cette étude concerne des gènes contrôlés par *mucA*, *lasR*, et *rpoN*. Ces trois mutations ont été retrouvées chez plus de 40 patients après 1979, ce qui suggère que toutes les souches correspondantes dérivent d'un ancêtre commun. Ceci pourrait également suggérer que la transmission massive de ces souches, retrouvées après 1979 chez les patients de la présente étude, mais également d'autres patients danois, est postérieure à l'apparition de ces trois mutations chez *P. aeruginosa* (Yang et al 2011).

La dispersion de cette souche clinique ancestrale et l'infection indépendante de plusieurs patients permettent également de rechercher une potentielle évolution parallèle des souches dérivées indépendamment chez les différents patients. Effectivement, un tel parallélisme a été mis en évidence, notamment pour les isolats issus des patients appelés CF173 et CF333, aussi bien au niveau génétique puisque plusieurs loci identiques sont mutés indépendamment chez ces deux types d'isolats, qu'au niveau phénotypique (Figure 9). En effet, la croissance avec l'acide 4-hydroxyphénylacétique (4-HPA) comme source de carbone et la résistance à la ciprofloxacine sont altérées de façon identique chez les deux types d'isolats et sont reliées à des mutations dans les mêmes gènes (Figure 9). Les interactions avec l'hôte imposent ainsi des pressions de sélection similaires sur les capacités adaptatives de bactéries. Un tel parallélisme a également été observé dans une étude similaire (mais

sur un temps plus court), au cours de laquelle 17 isolats cliniques de *P. aeruginosa* ont été collectés à partir de 3 patients atteints de mucoviscidose sur une période allant de 3 mois à 8 ans (Huse et al 2010). Les profils de transcription globaux de l'expression des gènes ont en effet révélé plus de 20 gènes dont la régulation a été altérée de façon commune chez ces isolats bactériens, dont certains importants pour la croissance *in vivo*.

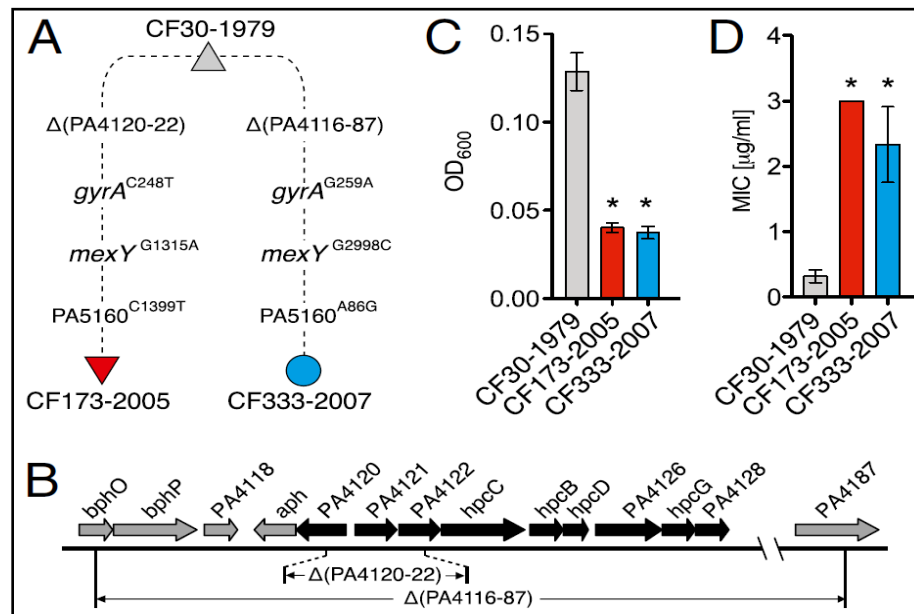


Figure 9 : Evolution parallèle dans les isolats CF173 et CF333 (d'après Yang et al 2011). (A) Représentation schématique de quatre changements génétiques parallèles et indépendants observés chez les clones CF173 (en rouge) et CF333 (en bleu) issus de deux patients différents. Les changements génétiques incluent des mutations ponctuelles dans trois gènes (*gyrA*, *mexY*, et *PA5160*) et des délétions des gènes impliqués dans le catabolisme du 4-HPA. (B) Schéma de la région chromosomique portant les gènes requis dans le catabolisme du 4-HPA (indiqués en noir) ainsi que les positions des délétions $\Delta(PA4120-22)$ (CF173) et $\Delta(PA4116-87)$ (CF333). (C) Croissance sur un milieu minimum avec du 4-HPA comme seule source de carbone. (D) Valeurs des concentrations minimales inhibitrices en ciprofloxacine (MIC; µg/mL).

Une étude analogue a été menée sur une durée de 16 années à Boston (Massachusetts, USA) lors de l'émergence d'une épidémie due à la bactérie

Burkholderia dolosa chez 14 patients atteints de mucoviscidose, dont le tout premier à avoir été infecté par cette bactérie (Lieberman et al 2011). La plupart des 112 isolats collectés de *B. dolosa* ont été obtenus à partir des voies aériennes, quelques-uns provenant cependant du sang dans les cas de septicémie (Figure 10).

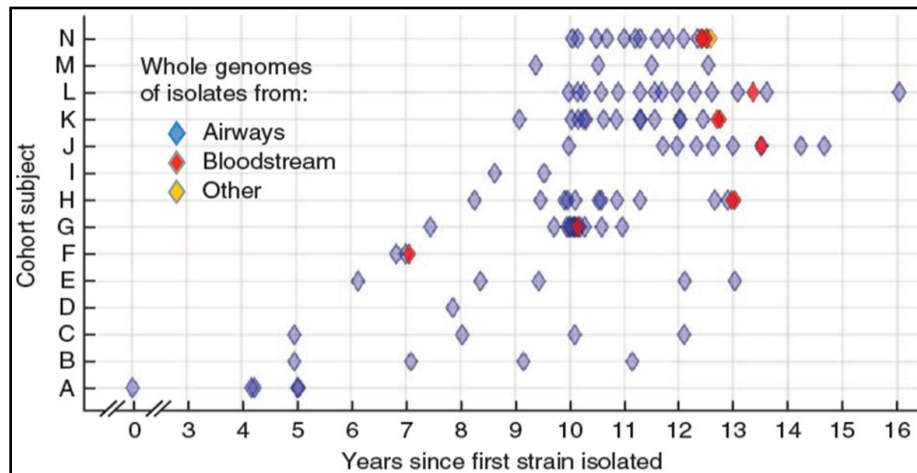


Figure 10 : Prélèvement de 112 isolats épidémiques de *B. dolosa* auprès de 14 patients atteints de mucoviscidose sur une période de 16 ans (d'après Lieberman et al 2011). Le génome des 112 isolats bactériens a été séquencé (losanges, chaque ligne horizontale correspondant à un patient). Les isolats ont été obtenus au cours du temps à partir des voies respiratoires (en bleu), de la circulation sanguine (en rouge), ou d'autres parties du corps (en jaune; par exemple, de tissus obtenus lors d'actes chirurgicaux) des patients.

Le génome des 112 isolats a été séquencé, prodiguant ainsi une analyse fine de l'évolution parallèle d'une même souche chez de multiples patients. Les résultats montrent que tous les isolats dérivent de la souche ayant infecté le patient initial et révèlent une diversification importante au cours de l'épidémie (~2,1 mutations ponctuelles par an). Les isolats issus d'un même patient tendent à former des groupes génétiquement liés issus d'une contamination unique suivie d'une diversification (Figure 11a). Néanmoins, les premiers patients infectés semblent avoir été contaminés à plusieurs reprises, et par des clones distincts (patients B, C, E, F). La comparaison de

ces groupes entre eux a permis de reconstruire l'histoire de la transmission du pathogène à partir du premier patient infecté (Figure 11b). La comparaison des isolats pulmonaires et sanguins chez un même patient suggère une contamination d'un organe à l'autre soit par une unique transmission multi-clonale soit par de multiples transmissions impliquant un clone unique (Figure 11c). En effet, quand des isolats multiples, collectés à partir du sang d'un patient, ont été analysés, ceux-ci dérivent de lignées différentes présentes dans le poumon de ce même patient. Il est à noter que des événements de diversification bactérienne se sont produits chez plusieurs patients (Figure 11), conduisant à des sous-lignées qui co-existent (patients G et L) et qui peuvent occasionner une septicémie imputée à plusieurs de ces sous-lignées (patients K et N). Ces événements de diversification sont probablement rendus possibles par la structuration spatiale et nutritionnelle extrêmement complexe, bien qu'assez méconnue, que revêt le tissu pulmonaire dans le cas de la mucoviscidose (mucus sécrété par les cellules épithéliales + biofilm fabriqué par les bactéries) (Hogardt et Heesemann 2010). D'autre part, les auteurs ont recherché les mutations permettant l'adaptation des bactéries à leur hôte au cours de l'infection, afin de comprendre l'évolution de ce pathogène. Les analyses génomiques n'ont révélé aucune signature de sélection dans les régions codantes ($dN/dS = \sim 1$). Afin d'identifier les gènes soumis à la sélection naturelle, ceux possédant plusieurs mutations (>3) ont été recherchés. Les 17 gènes ainsi identifiés ont, quant à eux, révélé une très forte signature de sélection positive ($dN/dS = 18$). La plupart de ces gènes (11/17) sont impliqués dans la pathogénie (synthèse membranaire, sécrétion, résistance aux antibiotiques). Cinq des 6 autres sont impliqués dans la régulation de l'expression des gènes (méthyltransférase, facteur

sigma, facteurs de régulation liés à la concentration de l'oxygène). Ces 17 gènes ont acquis de multiples mutations non-synonymes et ce, chez de nombreux isolats bactériens collectés dans différents patients, ce qui révèle à nouveau un fort degré de parallélisme.

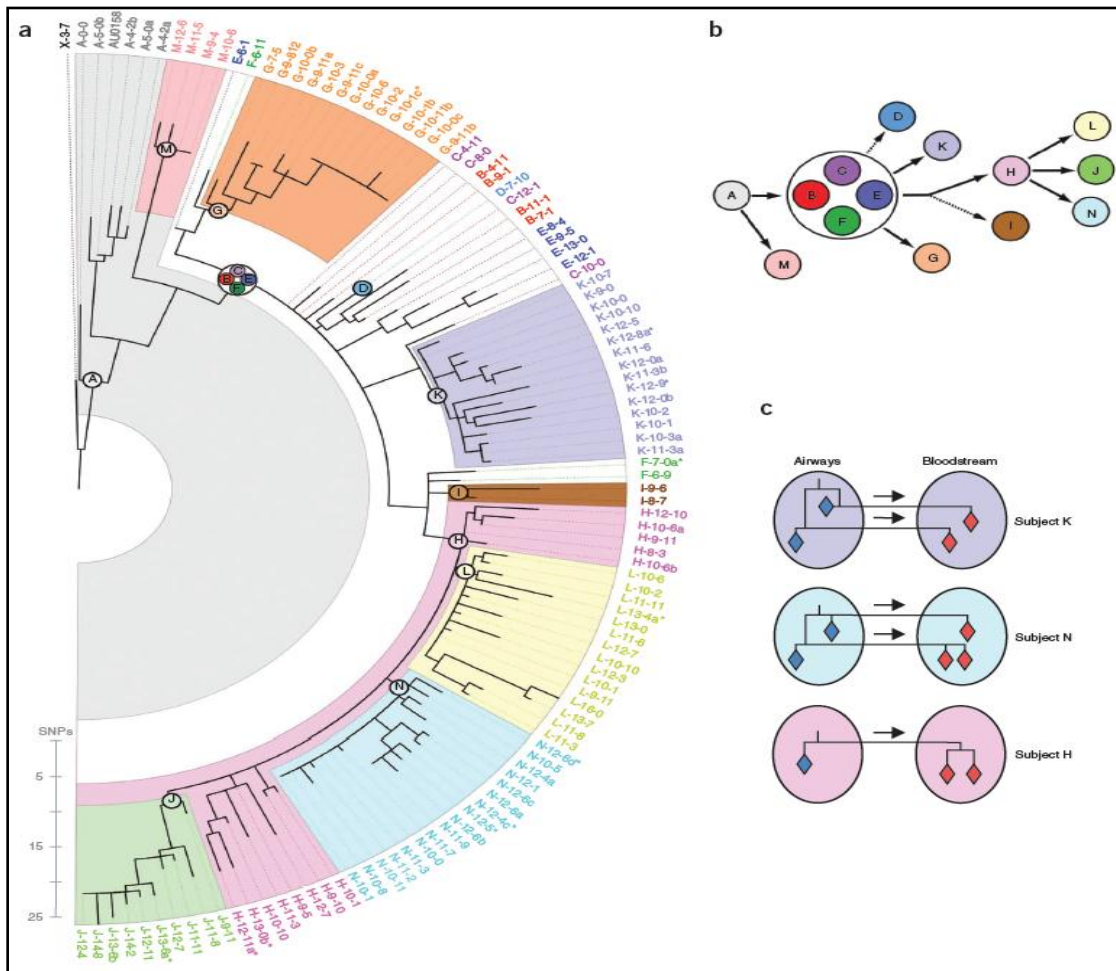


Figure 11 : La phylogénie bactérienne révèle un réseau probable de transmission entre individus et entre organes (d'après Lieberman et al 2011). (a) Arbre phylogénétique basé sur le nombre de SNPs. Les 112 isolats sont indiqués par les fines lignes pointillées et colorées en fonction du patient. Le plus ancien ancêtre commun des isolats d'un même patient est représenté par un cercle avec le sigle et la couleur appropriés. Les fonds colorés indiquent une empreinte génétique spécifique à un patient. Par exemple, les patients B, C, E et F partagent le même ancêtre commun (fond blanc). (b) La phylogénie entre les différents ancêtres communs suggère un réseau probable de transmission de l'infection entre patients (flèches). Les flèches pointillées indiquent une certitude moindre (moins de trois isolats). (c) La phylogénie entre les isolats issus du sang et des poumons prélevés sur un même patient

montre une transmission de multiples clones vers la circulation sanguine au cours de septicémies (flèches multiples, patients K et N).

II. Les événements de diversification dans des environnements expérimentaux

Les stratégies d'évolution expérimentale semblent constituer un système idéal pour étudier des événements évolutifs tels que les processus de diversification. En effet, la multiplication rapide et la grande taille de population des microorganismes utilisés permettent d'obtenir facilement un grand nombre de générations, et leur capacité à être congelés puis revivifiés dans l'état où ils ont été prélevés permet de remonter le temps et donc d'étudier des processus évolutifs « en temps réel ». En analysant les intermédiaires évolutifs conservés à partir de telles populations expérimentales, il est donc en théorie possible de décortiquer les étapes nécessaires à la mise en place des processus de diversification, en particulier dans le cas de polymorphismes stables. En réalité, de tels événements d'émergence et de maintien de polymorphismes restent rares, y compris en laboratoire, où la diversité génétique et phénotypique observée au sein d'une population est souvent transitoire. En effet, elle relève davantage d'événements successifs de sélection périodique, associés ou non à des processus d'interférence clonale entre des sous-populations portant des mutations ayant un effet similaire sur la valeur sélective et co-existant à un moment évolutif donné. La littérature recense cependant des cas de diversification adaptative dans des lignées d'évolution expérimentale, souvent apparus dans des environnements hétérogènes (structuration spatiale, présence de plusieurs ressources nutritives disponibles). En effet, la structuration du milieu par l'absence d'agitation et donc différents degrés d'oxygénation

(Rainey et Travisano 1998) ou par la présence de plusieurs sources de carbone (Spencer et al 2007) crée différentes niches écologiques, ce qui est propice à l'apparition d'une diversité liée à la spécialisation cellulaire. Cependant, une diversification peut également être mise en évidence même dans des environnements homogènes, qui peuvent être rendus hétérogènes non pas par l'expérimentateur, mais par les bactéries elles-mêmes. Ce phénomène est appelé construction de niches. En effet, la capacité de cellules à sécréter des composés peut dans certains cas induire une spécialisation d'une partie de la population pour ces sources de carbone initialement absentes du milieu de l'expérience (Rosenzweig et al 1994). Une observation plus poussée de ces différents cas décrits de diversification adaptative fait ressortir un point commun entre eux dans chacune des conditions spécifiques où ces diversifications ont été analysées : leur apparition répétée (parallèle) pour différents réplicats propagés à partir d'un ancêtre commun dans des environnements similaires. De plus, et dans chacun de ces cas, les analyses génétiques disponibles montrent que l'événement de diversification est associé à des changements mutationnels uniques, ce qui est probablement à la base de ce parallélisme phénotypique.

1. Environnements complexes générés par construction de niche

Des populations indépendantes d'*E. coli* ont été propagées à partir d'un clone ancestral génétiquement homogène dans des conditions de croissance continue en chimostat, avec comme seule source de carbone du glucose en quantité limitante (0,05 - 0,1%). Après seulement 765 générations, et durant les 450 générations ultérieures analysées,

trois types cellulaires phénotypiquement différenciés (sur la base de la taille des colonies générées après étalements sur boîtes de Pétri) ont été identifiés et co-existent (Rosenzweig et al 1994). Ces trois types cellulaires peuvent être cultivés séparément. Des co-cultures ont été réalisées en mélangeant soit un clone de chacun des trois types, nommés respectivement CV101, CV103 et CV116, soit un clone de deux des types cellulaires. Ces co-cultures ont montré que les trois types de clones étaient capables de co-exister soit tous ensemble (Figure 12a), soit deux à deux dans chacune des trois combinaisons possibles (Figure 12b, c, d).

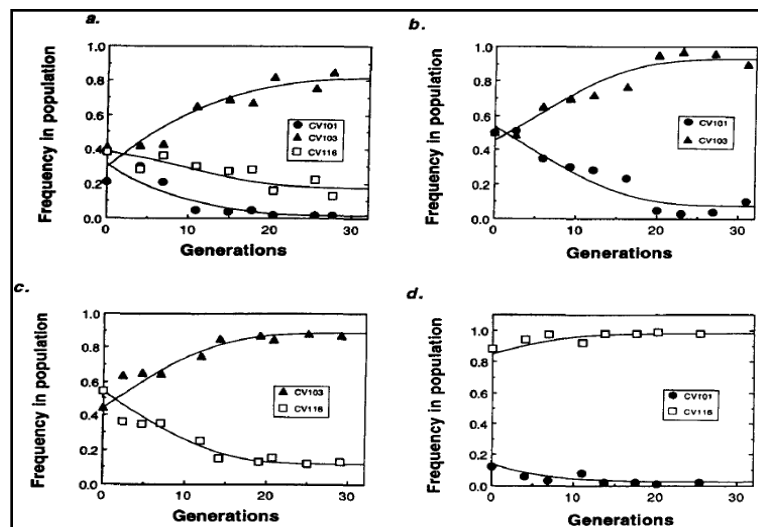


Figure 12 : Expériences de co-cultures démontrant que les trois clones CV101, CV103 et CV116 constituent un polymorphisme stable (d'après Rosenzweig et al 1994). (a) Compétition entre les clones CV101, CV103 et CV116. (b) Compétition entre les clones CV101 et CV103. (c) Compétition entre les clones CV103 et CV116. (d) Compétition entre les clones CV101 et CV116.

Une analyse approfondie de ces trois types de clones évolués et de leurs interactions a révélé que leur co-existence pouvait s'expliquer par des capacités différentielles de sécrétion et de consommation de deux composés différents, l'acétate et le glycérol (Rosenzweig et al 1994). Ainsi le clone CV103, isolé de la sous-population majoritaire,

possède deux phénotypes spécifiques majeurs : une capacité meilleure à assimiler le glucose, ce qui représente un avantage sélectif au sein du chemostat où celui-ci est présent en quantité limitante, et la capacité nouvelle de sécréter de l'acétate et du glycérol, cette dernière résultant probablement de réorganisations métaboliques suite à la meilleure utilisation du glucose. L'apparition de ce clone, capable de sécréter des métabolites dans le milieu extracellulaire, a conduit à la création de nouvelles niches écologiques qui ont alors permis l'émergence des deux autres types cellulaires. Le clone CV101 semble avoir acquis la capacité constitutive à consommer l'acétate, alors que celle-ci est normalement inductible et soumise à la répression catabolique par la présence de glucose (Brown et al 1977). Enfin, le clone CV116 montre une meilleure consommation du glycérol dans les conditions de l'évolution expérimentale en chémostat. Ainsi, la diversification des ressources disponibles dans le milieu et donc la construction de nouvelles niches écologiques par l'activité métabolique des bactéries elles-mêmes (Odling-Smee et al 1996, Laland et al 1999, Kassen et Rainey 2004) ont rendu possible l'émergence d'un polymorphisme complexe au sein d'un environnement initialement homogène et non structuré.

Afin de déterminer les bases génétiques de ce polymorphisme, l'expérience a été répétée douze fois en parallèle, dans strictement les mêmes conditions (Treves et al 1998). Cinq des douze populations indépendantes présentent, après quelques centaines de générations d'évolution, un polymorphisme lié à l'apparition de clones capables d'utiliser l'acétate sécrété dans le milieu par une autre partie de la population. Dans ces clones utilisateurs d'acétate, l'activité spécifique de l'enzyme acetyl-CoA-synthetase (ACS), impliquée dans la consommation d'acétate, est augmentée de plus

de 50 fois en présence de glucose par rapport à l'ancêtre, alors que le gène *acs* est normalement soumis à la répression catabolique et donc non exprimé en présence de glucose. L'expression de *acs* est donc non seulement accrue mais également constitutive, ce qui suggère un changement de régulation. En effet, l'expression constitutive accrue de *acs* a pour base génétique des modifications dans la région promotrice du gène (Figure 13), dues à des insertions d'éléments transposables (IS30 ou IS3) ou à une mutation ponctuelle (T-93A). Les insertions d'IS produisent ici de nouveaux promoteurs pour *acs* qui sont non régulés par la répression catabolique et donc constitutifs, alors que la mutation ponctuelle génère une augmentation constitutive de la transcription. Il est important de constater que les mutations à la base de cet événement de diversification correspondent à des changements de régulation et non pas à des modifications structurales des enzymes impliquées.

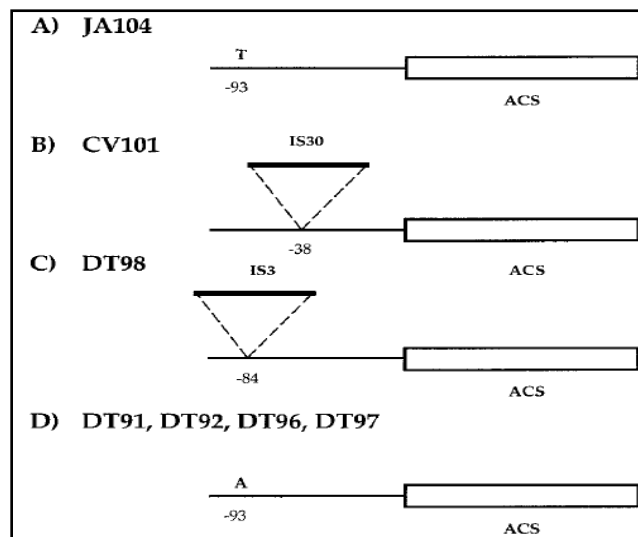


Figure 13 : Structure du gène *acs* chez les clones d'*E. coli* ancestraux et évolués consommateurs d'acétate (d'après Rosenzweig et al 1994). (A) Souche ancestrale JA104. (B) Insertion d'IS30 dans la région régulatrice du gène *acs* dans le clone consommateur d'acétate CV101. (C) Insertion d'IS3 dans la région régulatrice du gène *acs* dans les clones consommateurs d'acétate issus d'une population d'évolution. (D) Mutation T-93A dans la région régulatrice du gène *acs* dans les clones consommateurs d'acétate issus de quatre populations d'évolution.

2. Environnements complexes générés par structuration spatiale

a. Ecologie

Pseudomonas fluorescens est une bactérie aérobie isolée dans le sol, dans l'eau et en interactions avec des plantes. Des microcosmes ont été développés en plaçant un clone ancestral de cette bactérie, formant des colonies lisses (SM, Smooth Morph) après étalement sur boîtes de Pétri, dans un milieu liquide défini sans agitation (Rainey et Travisano 1998). Ce système expérimental produit une hétérogénéité spatiale liée notamment à un gradient de concentration en oxygène, élevée à l'interface air-liquide puis diminuant jusqu'au fond où des conditions anaérobies prédominent. Très rapidement, après sept jours dans ces conditions et dans la totalité des lignées indépendantes initiées, une importante diversité phénotypique a été observée après étalement sur milieu solide. Trois types de colonies correspondant à trois classes phénotypiques majeures ont été observées (Figure 14a et b) : 1. SM (Smooth Morph) d'aspect ancestral lisse ; 2. WS (Wrinkly Spreader) formant des colonies étalées et ridées ; 3. FS (Fuzzy Spreader) formant des colonies étalées et crépues. Quand chacun de ces trois types cellulaires est inoculé dans le microcosme de départ, chacun colonise une niche écologique différente : SM colonise le milieu liquide dans son intégralité, WS colonise spécifiquement l'interface air-liquide en fabriquant une structure de type biofilm, et FS colonise spécifiquement le fond du microcosme en milieu liquide (Figure 14b et c). Cet événement de radiation est donc strictement lié à la présence de niches écologiques différentes, générant des opportunités pour chacun des types cellulaires issus du clone ancestral.

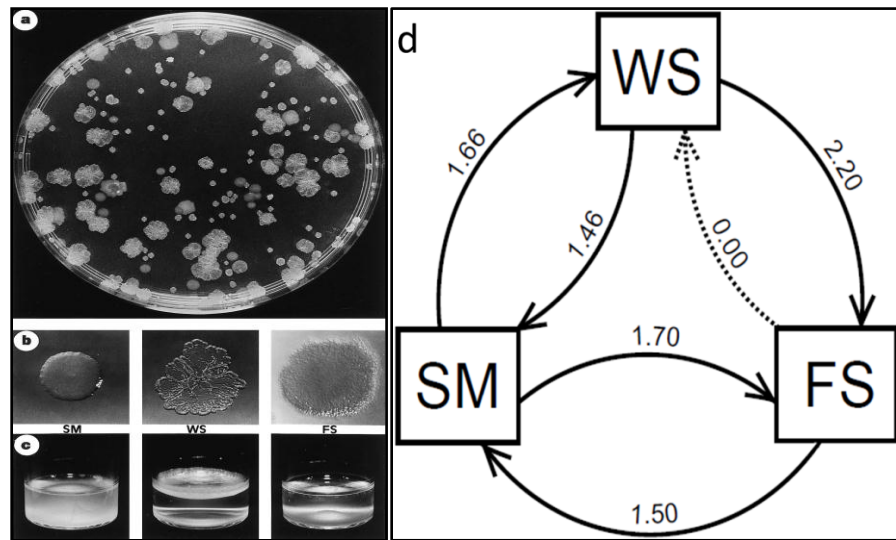


Figure 14 : Diversité phénotypique, spécificité de niche et relations écologiques entre les cellules de *P. fluorescens* ayant évolué dans un environnement spatialement hétérogène (d'après Rainey et Travisano 1998). (a) Après 7 jours, les populations révèlent une diversité phénotypique importante visualisée après étalement sur gélose. (b) La plupart des variants phénotypiques peuvent être classés dans l'un des trois types suivants: Smooth Morph (SM), Wrinkly Spreader (WS) et Fuzzy Spreader (FS). (c) Chacun des trois types cellulaires colonise spécifiquement une niche écologique. (d) Les trois types cellulaires interagissent par des relations dépendantes de la fréquence. Dans cinq des six cas, quand un type cellulaire est rare, il envahit le type cellulaire fréquent. Les flèches pointent du génotype initialement rare vers le génotype initialement fréquent pour chaque compétition. Les valeurs indiquées correspondent à la valeur sélective moyenne du clone initialement rare.

Ceci est d'autant plus évident que cette radiation n'a été observée que si l'expérimentateur permet la présence de ces niches écologiques. En effet, la diversification ne peut pas s'établir lorsque l'environnement est homogénéisé par agitation. De plus, une fois la radiation mise en place, une agitation de l'environnement la fait disparaître. Ainsi, l'apparition et le maintien de chacun des types cellulaires ne sont possibles que par la structuration spatiale de l'environnement et l'opportunité

d'accéder aux différentes niches écologiques qui résultent de cette structuration. De plus, chacun des nouveaux types cellulaires WS et FS présente des interactions négatives dépendantes de la fréquence avec le type SM, apparenté à la forme ancestrale (Figure 14d).

Ainsi, la co-existence entre ces trois types cellulaires résulte de la combinaison de deux phénomènes : les interactions négatives dépendantes de la fréquence et les compromis évolutifs liés à la spécificité de chaque type pour une niche écologique définie. Les interactions écologiques ont surtout été étudiées dans le cas du type cellulaire WS par rapport au type ancestral SM.

b. Génétique de la radiation adaptative

Bien que les bases génétiques des formes évoluées SM et FS n'aient pas été étudiées jusqu'à maintenant, de nombreuses analyses moléculaires ont été menées sur le type cellulaire WS, qui présente l'innovation phénotypique la plus remarquable avec la capacité nouvelle de former une structure de type biofilm permettant la colonisation de l'interface air-liquide. Une mutagenèse par transposon d'un clone évolué WS a permis de mettre en évidence l'implication de l'opéron *wss* (Figure 15) dans le phénotype WS (Spiers et al 2002). Cet opéron code dix protéines impliquées dans la production d'un polymère apparenté à la cellulose et n'est pas exprimé chez l'ancêtre. C'est la surproduction de la forme acétylée de ce polymère qui donne aux WS leur phénotype, leur permet de former une structure de type biofilm et leur confère un avantage adaptatif dans ces conditions (Spiers et al 2002, 2003).

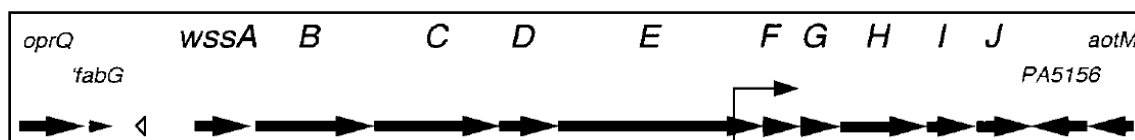


Figure 15 : Structure de l'opéron wss (d'après Spiers et al 2002). L'opéron wss contient 10 gènes permettant potentiellement la formation d'un polymère cellulosique.

Malgré l'implication de cet opéron dans la capacité des WS à coloniser l'interface air-liquide, aucune mutation n'y a été identifiée chez des clones évolués WS. Ainsi, des voies régulatrices de cet opéron pourraient être impliquées. Des travaux plus récents ont en effet montré qu'au moins trois voies de signalisation différentes, situées toutes trois en amont de la cascade de régulation qui amène à la production de cellulose, sont affectées par des mutations dans les clones évolués WS (Bantinaki et al 2007, McDonald et al 2009). Ces trois voies de signalisation ont en commun des enzymes de type di-guanylate cyclase (DGC), qui sont responsables de la biosynthèse de c-di-GMP, ce dernier étant l'activateur des enzymes de synthèse de la cellulose (Figure 16a) (Malone et al 2007). De plus, le c-di-GMP est connu pour constituer une molécule-signal chez un grand nombre de bactéries telles que *Acetobacter xylinum*, *Agrobacterium tumefaciens*, *Caulobacter crescentus*, *Vibrio cholerae*, *Salmonella typhimurium*, *Salmonella enteritidis*, *P. fluorescens*, *P. aeruginosa*, *Myxococcus xanthus* et *E. coli*, et est impliqué dans des processus de modulation des structures de surfaces des cellules bactériennes, et plus spécifiquement de la biosynthèse de cellulose chez ces organismes (Jenal 2004, voir Introduction Partie I section II.2).

La première de ces trois voies, nommée Wsp, contient un complexe protéique récepteur de signal lié à la membrane et constitué de quatre protéines (WspA, WspB, WspD et WspE), le signal étant donné par deux enzymes aux activités opposées (une

méthyltransférase nommée WspC et une méthylestérase nommée WspF). L'ensemble contrôle l'activité de la dernière enzyme, WspR, qui catalyse la synthèse de c-di-GMP (Figure 16a). Des mutations ponctuelles responsables du phénotype WS ont été trouvées dans les gènes régulateurs de la voie, *wspF* et *wspE* (Figure 16b). La deuxième voie, nommée Aws, comprend trois protéines, appelées AwsO, AwsR présentant une activité DGC, et AwsX qui régule l'activité de la protéine précédente (Figure 16a). Des mutations ont été identifiées dans les clones évolués WS dans chacun des trois gènes *awsO*, *awsR* et *awsX* (Figure 16b). Enfin la troisième voie, appelée Mws, contient une seule et unique enzyme connue à ce jour, MwsR, possédant une activité DGC et exerçant une autorégulation négative sur cette activité (Figure 16a). Des mutations ponctuelles responsables du phénotype WS ont été trouvées dans *mwsR* (Figure 16b). Il semble donc que les mutations à l'origine du phénotype WS touchent principalement des gènes codant des répresseurs de la synthèse de c-di-GMP, conduisant à sa surproduction et entraînant ensuite la surproduction de cellulose acétylée responsable du phénotype WS. Il est à noter que le génome de *P. fluorescens* contient 36 autres gènes codant potentiellement des enzymes possédant une activité DGC, ce qui offre encore beaucoup d'opportunités de mutations et d'évolution vers le phénotype WS.

Figure 16 : Détail des voies et loci affectés par des mutations dans les clones de phénotype WS (d'après McDonald et al 2009). (a) La surproduction de c-di-GMP résulte en une surproduction de cellulose et d'autres facteurs d'adhésion qui déterminent le phénotype WS. Le génome ancestral contient 39 DGCs potentielles, chacune étant en principe capable de synthétiser du c-di-GMP. Pour le moment, le génotype WS a été associé à des mutations dans trois des voies faisant intervenir des DGCs : Wsp, Aws et Mws. Dans chaque cas, les mutations trouvées le sont fréquemment dans les éléments de régulation négative : *wspF*, *awsX*, et dans le domaine phosphodiesterase de *mwsR*. (b) Les loci *wsp* et *aws* sont composés de sept et trois gènes respectivement; *mws* et *sws* sont composés de gènes uniques. Les gènes codant des DGCs dans les voies *wsp* et *aws*, ainsi que des domaines DGC dans les gènes *mws* et *sws* sont hachurés. PDE indique le domaine phosphodiesterase. Les flèches montrent la position des mutations responsables du phénotype WS dans les 26 mutants WS spontanés obtenus indépendamment dans cette étude. Les bases génétiques de l'activation de *swsR* sont inconnues.

Ainsi, cet événement de diversification, basé sur une structuration de l'environnement, a été rendu possible par des changements mutationnels de trois voies parallèles très différentes, mais qui ont en commun de contenir au moins un gène codant une protéine impliquée dans la production de la molécule signal c-di-GMP et soumise à régulation négative. Ces résultats soulignent à nouveau l'implication des réseaux de régulation dans l'évolution, l'adaptation et l'innovation phénotypique des bactéries.

c. Evolution de la coopération

L'apparition des WS et la production du polymère nécessaire à leur croissance à l'interface air-liquide (et donc l'accès à une nouvelle niche écologique) sont également caractérisées par une transition au niveau de l'unité sur laquelle porte la sélection naturelle. En effet, alors que chez les bactéries, qui sont unicellulaires et se divisent par fission binaire, celle-ci porte habituellement sur des individus, dans le cas de la radiation

adaptative chez *P. fluorescens* la sélection naturelle affecte non plus des cellules individuellement différenciées mais un groupe de cellules coopératrices et productrices d'un polymère qui constitue ici une ressource commune. Pour qu'un tel comportement coopérateur puisse persister dans le temps, le bénéfice de la coopération doit être supérieur au coût individuel de la production du polymère. Ce coût, estimé à une perte de 20% de la valeur sélective pour les types WS comparé à l'ancêtre, est compensé par la capacité à coloniser une nouvelle niche écologique, ici l'oxygène en surface, à laquelle le groupe a un accès privilégié (Rainey et Rainey 2003).

Dans le cas de coopérations, le coût au niveau des cellules individuelles d'un comportement coopératif favorise l'apparition de tricheurs (Figure 17), qui sont déficients pour la synthèse de la ressource commune (ici, la production du polymère), mais peuvent en profiter grâce à la présence des individus coopérateurs qui eux fabriquent la cellulose acétylée, maintenant le groupe à la surface du milieu de culture (Rainey et al 2000).

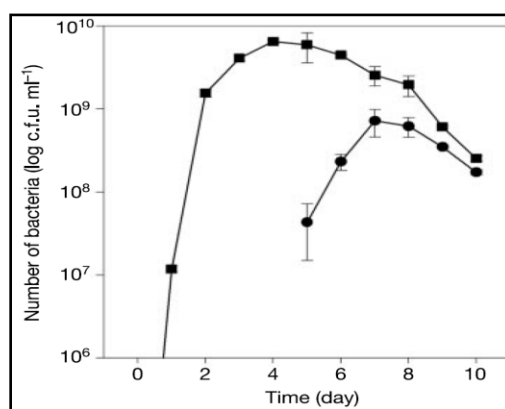


Figure 17 : Emergence de tricheurs à partir des clones WS au cours de l'évolution en environnement spatialement hétérogène (d'après Rainey et Rainey 2003). Les tricheurs apparaissent *de novo* par le biais de mutations et forment des colonies avec une morphologie lisse qui permet de les distinguer des clones WS après étalement sur boîtes. Les carrés représentent les clones WS et les cercles les tricheurs.

Quand des individus WS sont inoculés dans les microcosmes, des tricheurs apparaissent très rapidement (après 5 jours, Figure 17), et prolifèrent jusqu'à rendre le polymère trop fragile. Celui-ci finit alors par s'effondrer, et la majorité des cellules, coopérateurs et tricheurs, meurent (Figure 18; <http://evolution.massey.ac.nz/rainey/index.htm>). Très rarement, le polymère peut se reformer si des coopérateurs WS survivent et rejoignent la surface, signe que la sélection continue à agir au niveau de l'individu quand le trait coopératif est compromis (Rainey et Rainey 2003, <http://evolution.massey.ac.nz/rainey/index.htm>).



Figure 18 : Apparition de tricheurs et désintégration du polymère formé par les clones WS (d'après <http://evolution.massey.ac.nz/rainey/index.htm>). Un polymère intact est le produit des interactions coopératives de millions de cellules (photographie de gauche). Quand le polymère se fragilise suite à l'apparition de tricheurs, il coule au fond du tube (photographie du milieu). Quand on secoue brièvement le flacon, le polymère se casse en de multiples morceaux, montrant la fragilité acquise (photographie de droite).

Ainsi, l'apparition des WS et l'émergence de la coopération *de novo* sont des événements évolutifs majeurs, mais leur maintien au long terme est fragile de par la fréquence d'apparition des tricheurs. Ces résultats montrent cependant que la transition vers des ordres de complexité plus grande est accessible à la sélection naturelle, y compris en conditions de laboratoire.

D'autres expériences d'évolution en laboratoire ont également montré que des traits coopératifs pouvaient ré-émerger, après avoir été perdus (Velicer et Yu 2003). Cette fois, la coopération n'est pas liée à la synthèse de ressources communes, mais à la capacité de l'organisme étudié à se différencier morphologiquement. Ces expériences ont été réalisées chez une autre bactérie, *Myxococcus xanthus*, qui à la différence de *P. fluorescens*, est une bactérie naturellement coopérative. En effet, *M. xanthus* possède des comportements sociaux qui se manifestent par des déplacements en groupes de cellules sur des surfaces molles (« swarming ») pour acquérir de la nourriture, ou encore par la fabrication de corps de fructification (« fruiting bodies ») portant les spores qui permettront à la bactérie de survivre en cas de carence nutritionnelle par exemple. Lorsque *M. xanthus* perd ses comportements coopératifs (par délétion artificielle des gènes nécessaires à la motilité par exemple), il est possible de faire ré-émerger ces comportements par l'activation soit de voies parallèles à celles perdues précédemment (Velicer et Yu 2003), soit de nouvelles voies contrôlant le développement de la bactérie (Fiegna et al 2006, Yu et al 2010). Par exemple, une mutation a été identifiée dans un gène codant un petit ARN dont la fonction était jusqu'alors inconnue chez *M. xanthus*. L'analyse de cette mutation a permis de démontrer son implication dans la ré-émergence des comportements sociaux et d'identifier la fonction de ce sARN, qui est impliqué dans le contrôle du développement de *M. xanthus* (Yu et al 2010). La sélection naturelle a encore une fois ciblé un régulateur global (voir Introduction Partie I section IV).

Cependant, et contrairement à *P. fluorescens* où le phénomène de coopération qui apparaît en 7 jours *in situ* fait suite à un événement de radiation adaptative, ici,

l'émergence de ces nouveaux coopérateurs chez *M. xanthus* n'a pas été démontrée comme étant accompagnée de polymorphisme.

d. Evolution phénotypique

Récemment, Paul Rainey a utilisé le même type de microcosmes que ceux décrits ci-dessus, mais avec une variante importante. En effet, des lignées de *P. fluorescens* ont été initiées comme précédemment (Rainey et Travisano 1998), mais en alternant des cycles de croissances en milieu non agité (donc structuré) et agité (sans structuration spatiale). De plus, seuls des clones minoritaires et présentant des phénotypes différents de leur ancêtre immédiat ont été choisis pour initier le cycle de croissance suivant (Beaumont et al 2009). Cette stratégie a été déployée pour assurer une forte innovation phénotypique. Après seulement quatre cycles dans ces conditions, des génotypes ont été mis en évidence dans ces cultures avec un phénotype de type « variation de phase ». En effet, après étalements sur boîtes de Pétri, ces génotypes donnent naissance à un mélange de colonies présentant des aspects translucides et opaques (Figure 19). Quand elles sont repiquées, ces colonies conservent ce caractère polymorphe et instable. Des études par microscopie ont montré que les colonies opaques contiennent majoritairement des cellules encapsulées, alors que les colonies translucides contiennent majoritairement des cellules non-encapsulées. Ces génotypes ne se maintiennent pas si la pression de sélection est modifiée, ce qui indique que leur persistance est imputable à la combinaison des deux pressions de sélection utilisées, c'est-à-dire un environnement changeant et la sélection de clones minoritaires avant

chaque nouveau cycle de croissance (Beaumont et al 2009). Cette capacité à générer de la variabilité phénotypique a été comparée au phénomène de « bet-hedging » qui désigne la propension à prendre des risques, à faire des paris. En effet, parmi les variants générés, certains pourraient s'avérer potentiellement fructueux en terme évolutif et adaptatif.

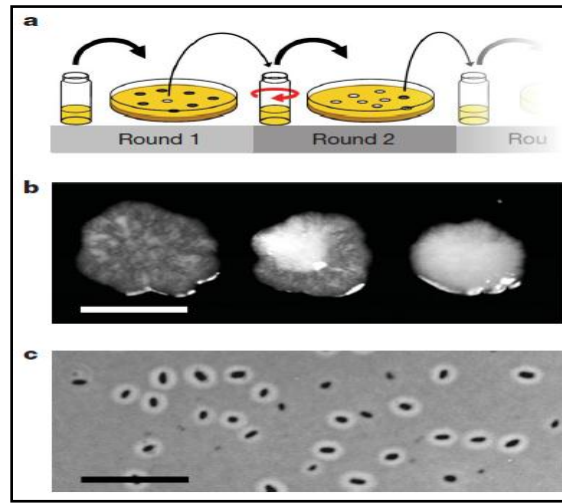


Figure 19 : Evolution de la morphologie des colonies (d'après Beaumont et al 2009). (a) Les populations ont été propagées dans un environnement cycliquement statique ou agité (flèche rouge), et criblées pour leur capacité à générer des cellules donnant des colonies d'un nouveau type, un prérequis pour le passage dans l'environnement suivant. La propagation entre chacun des environnements s'effectue par transfert des cellules provenant d'une seule colonie. (b) De gauche à droite : colonie translucide, mixte et opaque. Les colonies mixtes résultent d'une variation stochastique de l'expression de gènes impliqués dans la synthèse d'une capsule durant la formation de la colonie. La barre d'échelle est approximativement de 2mm. (c) Cellules capsulées et non-capsulées (microscopie à contraste de phase). La barre d'échelle est approximativement de 10 μ m.

La séquence du génome des génotypes présentant ce phénomène de bet-hedging a révélé les mutations spécifiques des WS (voir paragraphe II.2.b ci-dessus), ainsi qu'une mutation supplémentaire dans le gène *carB*, qui code la grande sous-unité de la carbamoylphosphate synthétase (CarAB), une enzyme centrale des voies de

biosynthèse des pyrimidines et de l'arginine (Mergeay et al 1974). Cette mutation, à elle seule, génère de la variation phénotypique quand elle est replacée dans le contexte génétique ancestral, ou la supprime quand elle est retirée du contexte évolué. Cette mutation n'est en revanche bénéfique que dans le contexte évolué, suggérant ainsi que la capacité de variabilité ne pourrait être bénéfique que dans de telles conditions et avec une telle sélection.

Ce phénomène de bet-hedging existe par ailleurs dans la nature : il s'agit des processus de variation de phase, dont par exemple la variation antigénique chez les parasites comme le trypanosome (Deitsch et al 2009), et de plasticité de notre système immunitaire pour la production des anticorps (Tonegawa 1983).

3. Environnements complexes générés par structuration nutritionnelle

Un clone d'*E. coli* a été utilisé comme ancêtre pour initier douze populations indépendantes qui ont été propagées par transferts quotidiens en culture discontinue dans un milieu minimum contenant du glucose et de l'acétate en quantités équivalentes (205 mg/L). L'utilisation de l'acétate est soumise à la répression catabolique chez *E. coli*, qui consomme donc tout d'abord le glucose jusqu'à quasi épuisement avant d'utiliser l'acétate comme source de carbone, ce qui est visualisé par une courbe de croissance de type diauxie. Après 1000 générations d'évolution dans ces conditions, deux types cellulaires ont été mis en évidence sur la base de la taille des colonies après étalements des cultures sur boîtes de Pétri. En effet, un mélange de petites et de grosses colonies sont visibles (Spencer et al 2007), chaque type de colonies présentant

un profil diauxique différent après inoculation en culture liquide dans les conditions de l'expérience d'évolution (Figure 20). Ainsi, les clones correspondant aux grosses colonies et appelés SS (Slow-Switcher, en rouge), présentent un profil de croissance quasi similaire à celui de l'ancêtre (en noir) avec tout d'abord une consommation du glucose puis une utilisation plutôt lente de l'acétate, alors que les clones correspondant aux petites colonies et appelés FS (Fast-Switcher, en bleu) présentent un profil de croissance avec un passage rapide de la consommation de glucose vers celle d'acétate et également une utilisation rapide de ce dernier. Les deux types cellulaires FS et SS co-existent grâce à des interactions négatives dépendantes de leur fréquence, les individus rares étant avantagés (Spencer et al 2007). Les différences phénotypiques entre les clones évolués FS et SS suggèrent à nouveau des changements génétiques au niveau de la régulation de l'utilisation des deux sources de carbone.

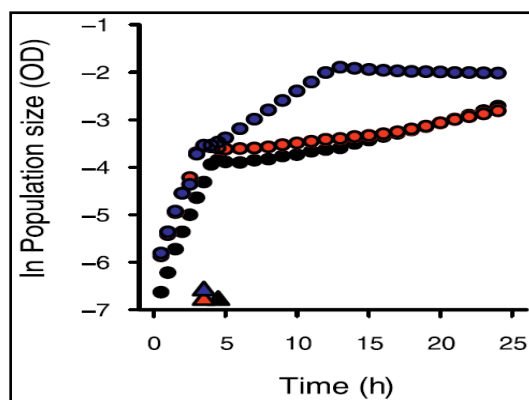


Figure 20 : Profils de croissance de chaque type cellulaire sur milieu glucose-acétate (d'après Spencer et al 2007). Les trois génotypes isolés (FS, SS, et ancêtre) consomment le glucose (entre 0 et 4 h), puis l'acétate (après 4 h), selon un profil de croissance diauxique. Les SS (en rouge) se rapprochent de la souche ancestrale (en noir) pour la croissance sur acétate, alors que les FS (en bleu) montrent un profil diauxique différent avec une transition plus rapide et un taux de croissance plus important sur acétate. Les triangles sur l'axe horizontal indiquent le moment de transition pour chaque souche.

La consommation d'acétate est dépendante de l'expression de l'opéron *aceBAK* (Figure 21), normalement soumis à la répression catabolique en présence de glucose. Les gènes *aceBAK* permettent la conversion de l'acétate et de ses dérivés en malate, qui est un intermédiaire du cycle de Krebs. Dans les clones FS, l'expression de ces gènes est dérégulée et constitutive, y compris en présence de glucose dans le milieu. Cette dérégulation s'explique par une insertion d'*IS 1* dans le gène *iclR* (Spencer et al 2007), qui code le répresseur de l'opéron *aceBAK* et qui est inactivé suite à cette insertion (Figure 21).

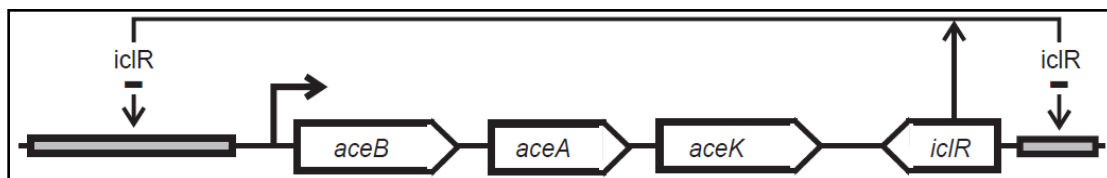


Figure 21 : Structure de l'opéron *aceBAK* et régulation négative par *IclR* (d'après Spencer et al 2007). Une défaillance de cette régulation négative, par exemple suite à l'inactivation du gène *iclR*, entraîne l'expression constitutive de l'opéron *aceBAK*.

Des profils globaux de transcription ont montré que cet événement de diversification était associé à des changements importants de l'expression des gènes impliqués dans le métabolisme central. Les clones évolués FS révèlent notamment une surexpression des gènes codant des enzymes impliquées dans le cycle de Krebs (*sucA*, *sucB*, *sucC*, *sucD* par exemple), le shunt du glyoxylate (*aceB*), la consommation d'acétate (*acs*) et la respiration anaérobie (*fadD*), alors que les clones évolués SS présentent une transcription accrue des gènes impliqués dans la sécrétion d'acétate (Figure 22). Ces résultats suggèrent ainsi une compétition entre les deux types cellulaires pour les sources de carbone et l'oxygène (Le Gac et al 2008).

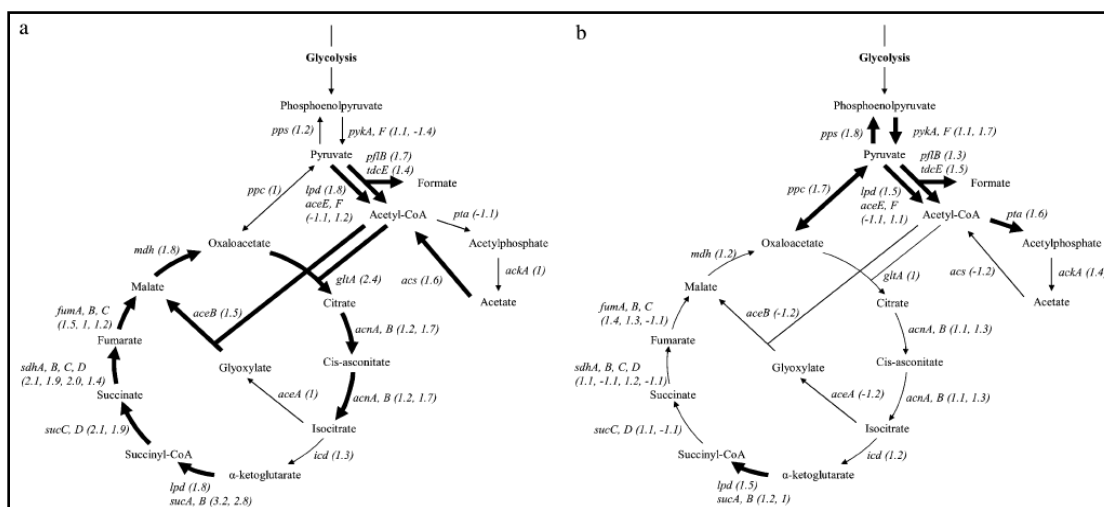


Figure 22 : Cycle de Krebs, shunt du glyoxylate, et voies métaboliques de consommation et/ou sécrétion d'acétate montrant des tendances d'expression différentes entre FS/ancêtre et SS/ancêtre (d'après Le Gac et al 2008). Les gènes codant les enzymes des voies métaboliques sont indiqués en italique et le facteur de changement d'expression est indiqué entre parenthèses. Une flèche épaisse indique un taux de changement $\geq 1,5$. **(a)** Comparaison entre FS et ancêtre. **(b)** Comparaison entre SS et ancêtre.

L'analyse de ces profils globaux de transcription par regroupement des gènes différentiellement exprimés en clusters de gènes co-régulés a révélé qu'un nombre important de gènes dont l'expression variait entre les deux écotypes FS et SS étaient contrôlés par ArcA, un régulateur global de la transcription de gènes dépendant des conditions d'oxygénation du milieu (Malpica et al 2004, voir Introduction Partie I section III). Une mutation a ainsi été retrouvée dans ce gène, spécifiquement chez les clones évolués FS (Le Gac et al 2008, 2010), validant ainsi les différences de transcription observées entre les deux types cellulaires. Ainsi, ce sont encore une fois les réseaux de régulation de l'expression des gènes, ici en liaison avec le métabolisme central, qui sont remaniés au cours de l'évolution pour conduire à l'émergence d'un polymorphisme.

INTRODUCTION PARTIE III

L'EXPERIENCE D'EVOLUTION AU

LONG TERME AVEC

ESCHERICHIA COLI

I. Le système expérimental

Parmi les expériences d'évolution microbienne développées pour examiner les processus adaptatifs et évolutifs, l'Expérience d'Evolution au Long Terme avec *E. coli*, que nous utilisons au laboratoire, est la plus longue en cours et également une des plus étudiées. Un clone ancestral d'*E. coli* B, appelé REL606 (Jeong et al 2009), a été utilisé pour initier douze populations indépendantes en février 1988 (Figure 23). Plus précisément, seules six populations, appelées Ara-1 à Ara-6, ont été initiées à partir de REL606, qui est par ailleurs incapable d'utiliser l'arabinose (Ara-) comme seule source de carbone (Lenski et al 1991). Les six autres populations, appelées Ara+1 à Ara+6, ont été initiées à partir d'un révertant Ara+ de REL606, appelé REL607. Le caractère d'utilisation de l'arabinose est utilisé comme marqueur phénotypique lors des expériences de compétition. Ce marqueur est neutre dans les conditions de l'évolution expérimentale (Lenski et al 1991). A l'exception d'une mutation ponctuelle neutre dans le gène *araA*, l'ensemble des douze populations est donc initialement génétiquement identique.

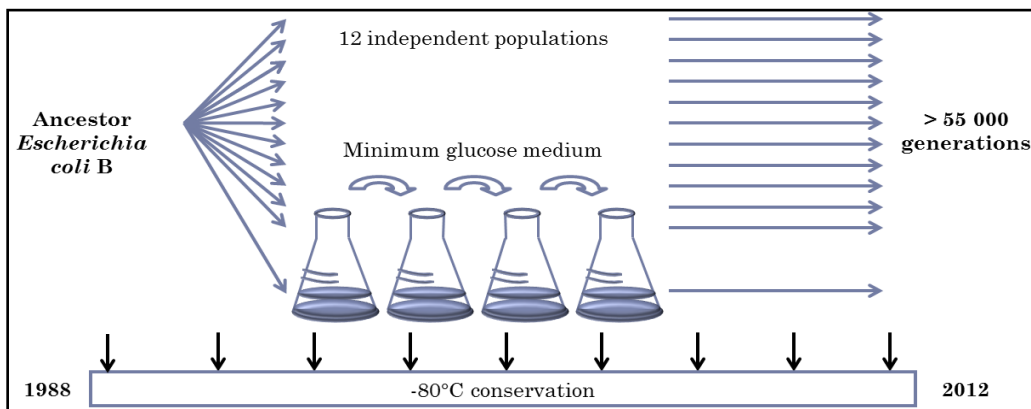


Figure 23 : Principe de l'expérience d'évolution au long terme. Un clone *Escherichia coli* B a été utilisé pour initier douze populations indépendantes, propagées quotidiennement depuis 1988 dans un milieu minimum contenant du

glucose en quantité limitante. Plus de 55 000 générations d'évolution ont ainsi été obtenues depuis le début de l'expérience, douze fois indépendamment. Des prélèvements ont été effectués toutes les 500 à 1000 générations, constituant ainsi de véritables archives fossiles.

L'ancêtre commun a été mis en culture dans douze flacons contenant chacun 10 mL de milieu minimum Davis (DM, Davis et Mingioli 1950), supplémenté avec une quantité limitante de glucose de 25 µg/mL (DM25), soit 139 µM. Chacune des douze populations a alors été incubée à 37°C avec agitation pendant 24h, puis propagée par transferts quotidiens de 100 µL dans 9,9 mL de milieu frais. Lors de chaque cycle de 24h, les populations expérimentent une phase de latence, suivie d'une phase de croissance et d'une phase stationnaire causée par l'épuisement du glucose disponible. Cet environnement se caractérise donc par un caractère saisonnier. Dans ces conditions, chaque population produit 6,64 générations durant chaque cycle quotidien ($\log_2 100$), conduisant à une taille de population de $\sim 5 \times 10^7$ cellules/mL après 24h d'incubation (Lenski et al 1991, Lenski 2004). Depuis le début de l'expérience en 1988, chacune des douze populations a ainsi évolué depuis plus de 55 000 générations à partir de leur ancêtre commun. Des échantillons de chaque population ont été prélevés et conservés à -80°C toutes les 500 à 1000 générations tout au long de l'expérience, ce qui permet de disposer de véritables archives fossiles revivifiables.

Les conditions de l'évolution se caractérisent ainsi par une extrême simplicité. De plus, le clone ancestral est strictement asexué : en effet, il n'est pas naturellement transformable et son génome ne contient aucun plasmide ou prophage actif (Jeong et al 2009, Studier et al 2009), éliminant ainsi toute possibilité évidente de transferts horizontaux de gènes. L'évolution au cours de cette expérience résulte donc de la

dérive génétique et de la sélection naturelle due aux variations générées par des mutations se produisant sur le chromosome (Lenski et al 1991, Lenski 2004). Les conditions simplifiées mises en œuvre dans cette expérience permettent ainsi d'appréhender la nature des processus évolutifs.

II. Parallélisme de l'évolution

Les douze populations ont montré une évolution parallèle de la valeur sélective tout au long de l'expérience. Elles se caractérisent par une forte augmentation de la valeur sélective par rapport à leur ancêtre commun au fil du temps évolutif, signe de leur adaptation à l'environnement. La trajectoire d'augmentation de la valeur sélective (Figure 24) est similaire pour les douze populations, avec un gain rapide durant les 2000 premières générations suivi d'une décélération importante tendant vers un plateau (Lenski et Travisano 1994, Lenski 2004, Philippe et al 2007).

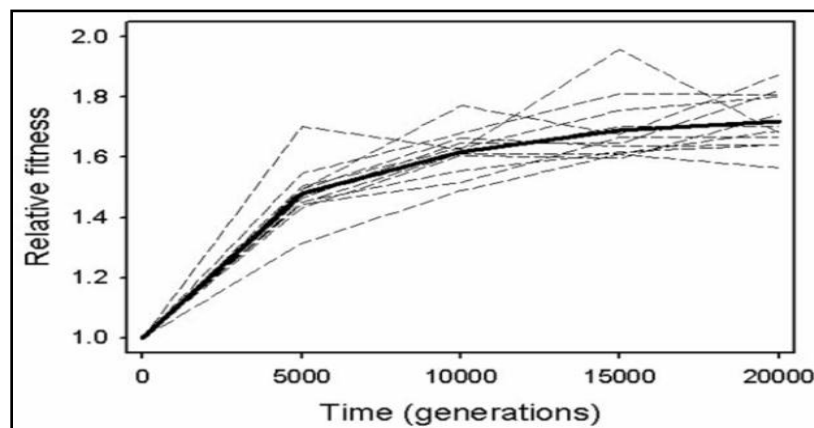


Figure 24 : Trajectoire de la valeur sélective au cours de 20 000 générations d'évolution expérimentale (d'après Philippe et al 2007). Chaque ligne pointillée représente l'évolution des valeurs de la valeur sélective de chaque population comparée à l'ancêtre. La ligne noire correspond à la moyenne de la valeur sélective des douze populations.

Associés à cette augmentation de la valeur sélective, d'autres phénotypes révèlent des changements parallèles dans la majorité des douze populations, dont la réduction du temps de latence après transfert dans du milieu frais, l'augmentation du taux de croissance maximal dans les conditions de l'expérience d'évolution et l'augmentation de la taille cellulaire (Lenski et Travisano 1994, Lenski 2004, Philippe et al 2009). De plus, toutes les populations ont montré une spécialisation catabolique au glucose, avec une perte partielle ou totale de croissance sur une variété de substrats alternatifs (Cooper et Lenski 2000). Dans le cas le plus drastique, les douze populations ont perdu la capacité de cataboliser le ribose (Cooper et al 2001). Les populations ont également expérimenté des changements parallèles de l'expression globale des gènes (Cooper et al 2003, Pelosi et al 2006), ainsi qu'un accroissement de la superhélicité négative de l'ADN (Crozat et al 2005, 2010). Ces derniers changements sont associés à des modifications des réseaux de régulation globale de l'expression des gènes (Philippe et al 2007, Hindré et al 2012). Enfin, la moitié des populations présentent un phénotype mutateur, et donc une augmentation du taux de mutation, après 55 000 générations (Sniegowski et al 1997, Lenski et al 2003, Barrick et al 2009).

Les analyses génétiques effectuées dans ces douze populations ont permis de montrer que les changements phénotypiques parallèles étaient souvent associés à des modifications génétiques dans des gènes similaires dans la majorité des populations. En effet, la perte de la capacité à utiliser des sources de carbone alternatives comme le ribose ou le maltose est associée à des délétions de l'opéron *rbs* dans les douze populations (Cooper et al 2001), et à des mutations dans le gène *malT* dans la plupart des populations (Pelosi et al 2006). Les modifications parallèles de taille cellulaire sont

quant à elles liées à des mutations dans le gène *pbpA*, impliqué dans la synthèse de la paroi cellulaire, dans la moitié des populations (Philippe et al 2009). D'autre part, les modifications de superhélicité de l'ADN sont associées à des mutations dans les gènes *topA*, *fis* et *dusB*, codant des topoisomérases et des régulateurs, et se produisant encore une fois dans la majorité des populations (Crozat et al 2005, 2010). Enfin, les modifications des profils globaux d'expression sont associées à des changements génétiques parallèles dans le gène *spoT* impliqué dans le contrôle des carences nutritionnelles (Cooper et al 2003, Pelosi et al 2006, voir Introduction Partie I section II.1), et à des modifications des interactions entre les régulateurs globaux de l'expression des gènes (Cooper et al 2008). Enfin, des mutations ont été mises en évidence dans la totalité des douze populations dans des gènes impliqués dans le métabolisme central comme *nadR* et *pykF*, qui sont impliqués dans la synthèse de NAD et la glycolyse, respectivement (Schneider et al 2000, Woods et al 2006). Etant donné que la grande majorité des gènes au sein des différentes populations non-mutatrices n'ont accumulé aucune mutation, ce parallélisme génétique reflète de façon probable l'action de la sélection naturelle (Lenski et al 2003). Ainsi, des expériences de reconstruction et de compétition ont montré qu'un grand nombre de ces mutations conféraient bien un avantage en terme de valeur sélective dans les conditions de l'expérience d'évolution (Cooper et al 2001, Pelosi et al 2006, Crozat et al 2005, Barrick et al 2009, Philippe et al 2009, Crozat et al 2010, Hindré et al 2012).

III. Evolution non parallèle et innovations phénotypiques

A côté de ce parallélisme remarquable, les douze populations présentent également des divergences évolutives. Alors qu'il existe un parallélisme au niveau des gènes cibles de la sélection naturelle, la position et la nature de ces changements diffèrent entre les populations, signe d'une divergence allélique (Cooper et al 2001, Cooper et al 2003, Pelosi et al 2006, Woods et al 2006, Stanek et al 2009, Crozat et al 2010). De plus, chaque population a accumulé des changements génomiques différents, notamment liés à l'activité des éléments génétiques mobiles IS (Papadopoulos et al 1999 ; Colin Raeside, Dominique Schneider et Joël Gaffé résultats non publiés). Si de telles divergences sont visibles entre les populations, il n'en reste pas moins qu'elles escaladent des pics adaptatifs qui peuvent être considérés comme similaires, dans les conditions simples et contrôlées de l'expérience d'évolution, où peu d'opportunités écologiques différentes sont disponibles. De plus, l'histoire évolutive même des populations pourrait restreindre, voire exclure, la découverte de telles opportunités.

Malgré ces contraintes, au moins deux des douze populations présentent une divergence écologique très importante, ainsi que des innovations phénotypiques majeures. Ces deux situations résultent de la découverte et de l'exploitation, par une partie des cellules constituant chacune de ces deux populations, d'opportunités écologiques qui ont été le résultat d'une histoire évolutive spécifique, un phénomène appelé contingence historique.

1. Capacité à utiliser le citrate

Dès le début de l'expérience, les conditions de l'environnement comportaient une opportunité écologique supplémentaire par rapport au glucose. En effet, le milieu DM25 dans lequel les douze populations sont propagées contient également 1,7 mM de citrate, en plus des 139 μ M de glucose. Cependant, cette abondante seconde source potentielle de carbone est inaccessible dans ces conditions car, malgré la présence d'un cycle de Krebs complet permettant le métabolisme interne du citrate durant la croissance en aérobiose sur d'autres substrats, *E. coli* est incapable de transporter et d'utiliser le citrate comme source de carbone dans des conditions oxygénées (Lara et Stokes 1952). Cette caractéristique est extrêmement stable et le phénotype Cit⁻ qui en résulte est depuis longtemps utilisé comme une des caractéristiques clé définissant l'espèce *E. coli* (Koser 1923, 1924, Scheutz et Strockbine 2005). Des isolats cliniques Cit⁺ d'*E. coli* ont été mis en évidence, mais ils résultent de transferts horizontaux de gènes (Ishiguro et al 1978, 1979). Un seul cas de mutant Cit⁺ spontané a été décrit à ce jour chez *E. coli* (Hall 1982). Ce mutant dérive d'une souche appelée D21 et a été isolé de façon totalement accidentelle lors d'une croissance sur un milieu contenant du phényle arabinoside comme source de carbone et du citrate comme agent chélateur du milieu. Les bases génétiques de ce phénotype n'ont pas pu être déterminées, mais l'auteur (Hall 1982) suggère que plusieurs mutations, ou des mutations complexes, sont à l'origine du phénotype mutant Cit⁺. Ainsi, *E. coli* possède le potentiel pour développer la capacité d'utiliser le citrate comme source de carbone dans des conditions aérobies, mais ce potentiel évolutif n'est pas facilement accessible et requiert des mutations

complexes et/ou multiples. Une des douze populations, appelée Ara-3, a fini par développer cette capacité après seize ans d'évolution, ce qui représente plus de 30 000 générations (Figure 25).

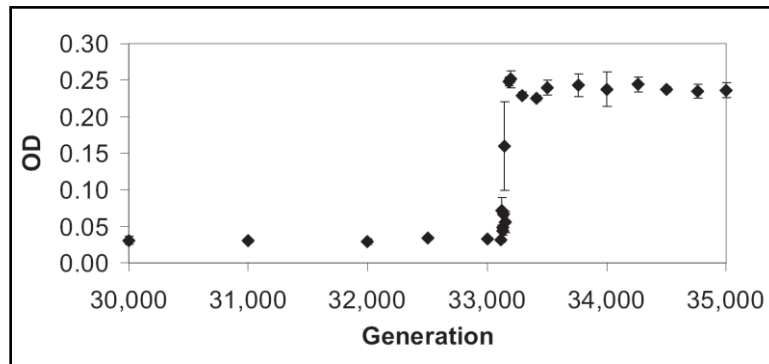


Figure 25 : Expansion de la population Ara-3 au cours de l'évolution du phénotype Cit+ (d'après Blount et al 2008). Des échantillons congelés à différents temps évolutifs de la population Ara-3 ont été revivifiés et utilisés pour initier des cultures en DM25. La densité optique à 420 nm (OD_{420}) a été mesurée pour chaque culture au bout de 24 heures.

Les archives fossiles de la population Ara-3, prélevées à différents temps au cours de l'évolution et conservées à -80°C , ont été utilisées pour rejouer le cours de l'évolution de cette population (Blount et al 2008). Ces expériences ont démontré que seules les populations, propagées à partir des prélèvements effectués après 20 000 générations d'évolution, étaient capables de produire des descendants Cit+, ce qui révèle que cette innovation phénotypique est strictement dépendante de l'histoire évolutive passée et donc de l'apparition antérieure d'une ou plusieurs mutations. Blount et al (2008) ont par ailleurs montré que la(es) mutation(s) préalable(s) potentialise(nt) l'émergence du phénotype Cit+ en augmentant le taux de mutation de l'ancêtre vers le phénotype Cit+ (Figure 26). Sans cette étape de potentialisation, le taux de mutation ancestral vers le phénotype Cit+ est incalculablement bas ($< 3,5 \times 10^{-13}$), alors que la présence de la(des)

mutation(s) potentialisante(s) rend cette fonction Cit⁺ accessible, même si le taux d'apparition reste faible ($1,7 \times 10^{-12}$).

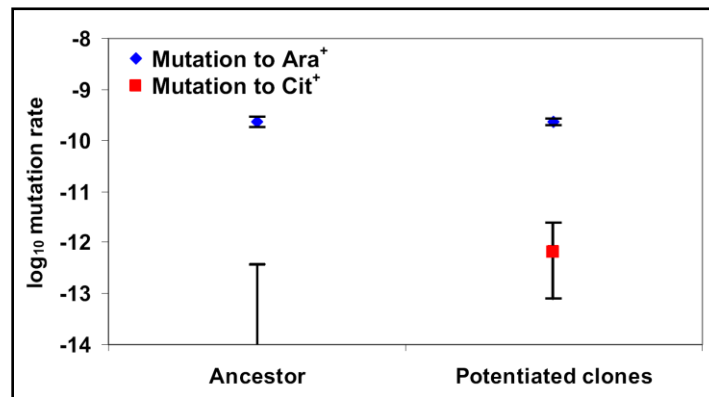


Figure 26 : La potentialisation permet d'augmenter le taux de mutation vers le phénotype Cit⁺ (d'après Blount et al 2008). Le taux de mutation de Ara⁻ vers Ara⁺, utilisé comme contrôle (losanges bleus), et de Cit⁻ vers Cit⁺ (losanges rouges) a été mesuré pour l'ancêtre et un mélange de clones potentialisants.

Cette population offre ainsi l'opportunité d'étudier la dynamique de cette innovation contingente de l'histoire évolutive passée. Le séquençage de génomes de plusieurs dizaines de clones, incluant des clones Cit⁺ et Cit⁻, isolés à de nombreux temps-clé de l'histoire de la population Ara-3 a permis de reconstruire une partie de son histoire évolutive (Blount et al 2012). La phylogénie, reconstruite à partir des séquences de génomes de ces clones, montre que la population est hétérogène pour la majeure partie de son histoire, avec trois clades Cit⁻ qui ont émergé vers 20 000 générations et qui co-existent jusqu'après l'émergence de l'innovation Cit⁺. Les clones Cit⁺ forment alors un clade monophylétique distinct qui émerge de l'un des trois clades Cit⁻ (Figure 27).

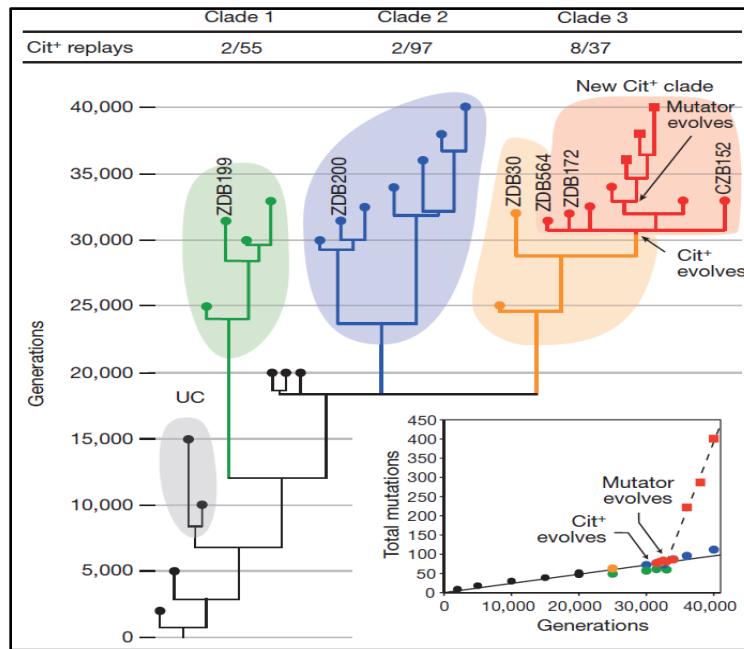


Figure 27 : Phylogénie de la population Ara-3 (d'après Blount et al 2012). Les symboles à chaque extrémité des branches montrent les 29 clones dont le génome a été séquencé. La couleur des symboles et des plages colorées correspond à chacun des clades majeurs; le clade UC (unsuccessful clade) en gris va à extinction après 15 000 générations; les clades 1, 2 et 3 en vert, bleu et orange respectivement composent la population Ara-3 entre 25 000 et 35 000 générations; puis les clones Cit+ en rouge émergent au sein du clade 3. Les symboles représentent un cercle quand le clone est non-mutateur, et un carré quand le clone est mutateur (la population devient mutatrice au sein de la lignée Cit+ vers 35 000 générations). Les rapports au-dessus de l'arbre montrent le nombre de clones appartenant à chaque clade qui ont abouti à un mutant Cit+ (Blount et al 2008) au cours des expériences pour rejouer le cours de l'évolution par rapport au nombre total de clones utilisés dans ces expériences. L'encadré montre le nombre de mutations par rapport à l'ancêtre. La ligne pleine montre la régression linéaire pour les génomes non mutateurs; la ligne pointillée pour les clones mutateurs.

La comparaison des séquences génomiques des clones Cit+ et Cit- a permis d'identifier la nature de la mutation permettant l'émergence de la fonction Cit+ (après la substitution d'une ou plusieurs mutation(s) potentialisante(s)) : il s'agit d'un événement de duplication en tandem qui conduit à la formation d'un nouveau module de régulation, au sein duquel le gène *citT*, codant le transporteur du citrate, est placé sous le contrôle

d'un nouveau promoteur qui permet son expression durant la croissance en aérobie (Figure 28). Cette duplication subit ensuite un événement d'amplification, conduisant à une expression encore plus élevée du transporteur, et donc à un phénotype Cit⁺ plus marqué (Blount et al 2012). Ces analyses génomiques n'ont cependant pas permis à ce jour d'identifier la(es) mutation(s) potentialisante(s), qui permet(tent) de favoriser l'accès à la fonction Cit⁺ obtenue via cette duplication en tandem.

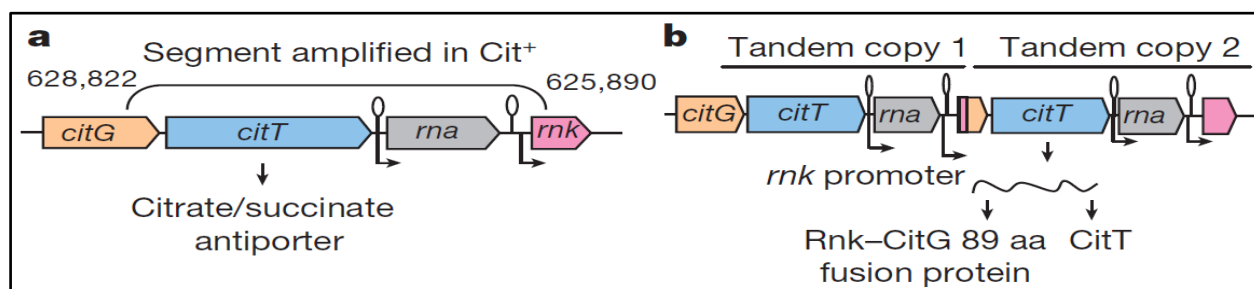


Figure 28 : Amplification en tandem dans les génomes des clones Cit⁺ (d'après Blount et al 2012). (a) Organisation ancestrale des gènes *citG*, *citT*, *rna* et *rnk*. (b) L'organisation spatiale et la régulation transcriptionnelle des gènes *rnk*, *citG* et *citT* sont altérées suite à l'amplification et amènent à la production d'une protéine fusion permettant l'utilisation du citrate en aérobie.

Au niveau écologique, l'émergence des clones Cit⁺ à 33 000 générations est associée à un événement de diversification adaptative à court terme, puisque les clones Cit⁻ persistent au sein de la population totale et présentent des interactions négatives dépendantes de la fréquence avec les clones Cit⁺ (Blount et al 2008). En effet, les clones Cit⁻ consomment plus efficacement le glucose, alors que les clones Cit⁺ se sont spécialisés pour le citrate. Ces derniers se caractérisent également par deux paramètres évolutifs extrêmement importants : d'une part, le citrate étant très abondant dans le milieu, les clones Cit⁺ sont majoritaires et représentent ~99% de la population totale, et d'autre part ils ont évolué un phénotype mutateur (Figure 27), ce qui n'est pas

le cas des clones Cit-. Ces deux paramètres pourraient contribuer à l'extinction des clones Cit-, qui a été observée vers 43 000 générations. La co-existence des clones Cit- et Cit+ est donc transitoire dans la population Ara-3 (C. Turner, Z. Blount et R. Lenski résultats non publiés).

2. Emergence d'un polymorphisme au long terme

Une des douze populations, appelée Ara-2, a évolué pour exploiter une opportunité écologique imprévue, produite en partie par une interaction que la conception même de l'expérience était supposée éliminer. La quantité de glucose présente dans le milieu DM25 est beaucoup plus faible que ce qui est typiquement ajouté aux milieux minimum de culture (Atlas 2004). L'utilisation d'une quantité aussi faible était liée, à l'origine, à l'objectif expérimental majeur d'examiner les dynamiques des relations phénotypes-génotypes en excluant le plus possible des niveaux élevés de complexité écologique (Lenski et al 1991, Lenski 2004, R. Lenski commentaires personnels). Réduire la quantité de glucose dans le milieu de culture permettait de réaliser cette simplification à deux niveaux : premièrement, en réduisant la taille maximale de la population, ce qui pouvait diminuer la probabilité d'interférence clonale, c'est-à-dire la co-existence de différentes sous-lignées cellulaires portant des mutations bénéfiques différentes et qui entrent ainsi en compétition au sein de la population et complexifient l'histoire évolutive (Gerrish et Lenski 1998, de Visser et Rozen 2006) ; deuxièmement, en restreignant la probabilité de sécrétion d'acétate et d'autres métabolites par les cellules et donc l'opportunité pour la mise en place d'un niveau de complexité écologique basé sur la

syntrophie et la construction de niche qui en découle (Helling et al 1987, Turner et al 1996). En effet, l'acétate est en général le métabolite majeur sécrété par les cellules d'*E. coli* quand elles se développent dans un milieu contenant du glucose en quantité élevée comme source de carbone. Malgré ces précautions, force est de constater la présence d'interférence clonale (Crozat et al 2005, Barrick et Lenski 2009, Woods et al 2011) et de diversification adaptative liée au phénomène de construction de niche (Rozen et al 2009, ce travail).

En effet, le potentiel d'interactions syntrophiques combiné à la saisonnalité de l'environnement expérimental, a ouvert une opportunité écologique que la population Ara-2 a réussi à exploiter. Celle-ci se caractérise par la présence simultanée de deux types cellulaires différenciés phénotypiquement, appelés S et L, en référence à la petite et grande taille des colonies qu'ils produisent respectivement après étalements sur boîtes de Pétri (Figure 29, Rozen et Lenski 2000). Les cellules donnant naissance aux petites colonies ont été détectées à partir de 6500 générations (Rozen et Lenski 2000) et les deux types cellulaires S et L co-existent maintenant depuis près de 50 000 générations (Le Gac et al 2012), et ce de manière dynamique (Figure 29). De plus, l'utilisation des séquences d'insertion IS comme marqueurs génomiques a permis de montrer que les lignées S et L constituaient des groupes monophylétiques, avec l'émergence de la lignée S lors de l'événement de divergence suivie de l'évolution et du maintien des deux lignées au cours de leur co-existence au long terme (Rozen et al 2005). En effet, plusieurs insertions d'IS permettent de distinguer génétiquement les clones appartenant aux deux lignées S et L.

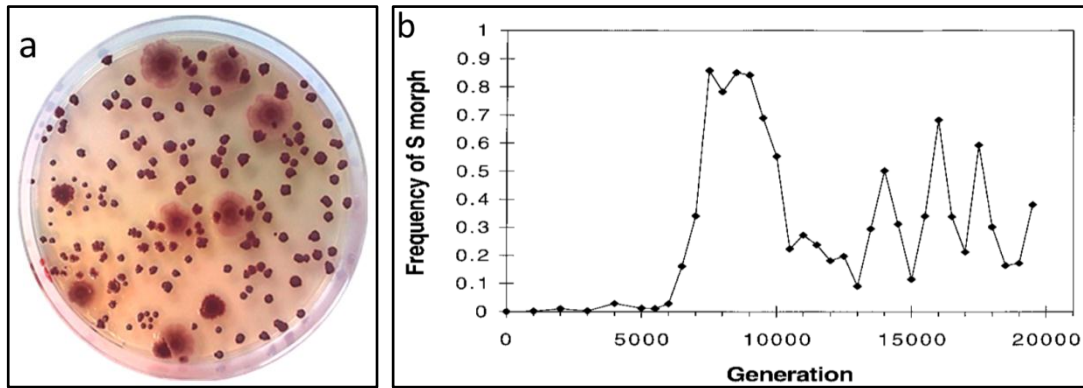


Figure 29 : Emergence et maintien du polymorphisme (d'après Rozen et Lenski 2000). (a) Morphologie des colonies après étalement sur gélose, les S forment de petites colonies et les L de grandes. (b) Chaque point testé montre la fréquence du type S, et a été obtenu en observant la taille des colonies après étalement des mélanges de la population Ara-2 sur boîtes de Pétri.

Cette co-existence au long terme, correspondant au polymorphisme le plus long et le plus stable jamais observé dans des conditions de laboratoire, est rendue possible grâce à des interactions négatives dépendantes de la fréquence, probablement basées sur une différenciation écologique entre les deux types cellulaires (Figure 30). Les lignées S et L peuvent ainsi être assimilées à des écotypes (Cohan 2008), et présentent en effet des caractéristiques phénotypiques et écologiques spécifiques.

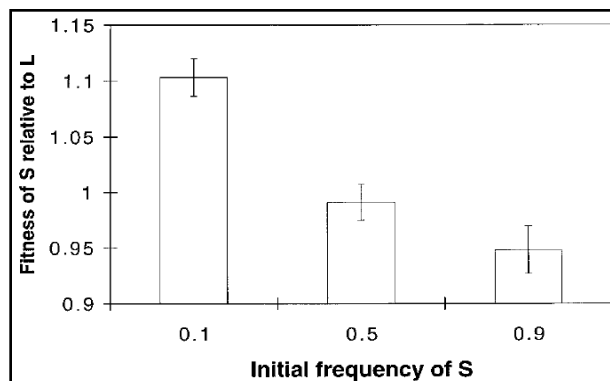


Figure 30 : Interactions négatives dépendantes de la fréquence entre les lignées S et L, mesurées par des expériences de compétition à court terme (d'après Rozen et Lenski 2000). La valeur sélective du génotype S par

rapport au génotype L est exprimée en fonction de la fréquence initiale des S. La valeur sélective est supérieure pour le génotype initialement rare.

Au sein de la population dans son ensemble, la lignée L présente un taux de croissance sur glucose supérieur de 20% à celui de la lignée S et un taux de mortalité inférieur en phase stationnaire (Figure 31). Cette dernière tendance est cependant inversée en présence des S, conditions dans lesquelles les L souffrent d'une mortalité accrue en phase stationnaire par rapport aux S (Figure 31). De plus, les cellules S sont capables de vivre aux dépens de ressources libérées par les cellules L (Rozen et al 2009). En effet, les S sont capables de se développer sur le surnageant issu de cultures de cellules L (Figure 32). Ainsi, la saisonnalité de l'environnement et les relations syntrophiques entre les deux écotypes ont procuré des niches écologiques différentes, que chacune des lignées S et L a exploitées, les cellules S et L étant plus adaptées à la phase stationnaire et exponentielle, respectivement (Rozen et al 2009). En effet, un accroissement artificiel de la durée de la phase exponentielle favorise la lignée L, alors que si la durée de la phase stationnaire est allongée, ce sont les S qui cette fois en bénéficient de façon privilégiée (Rozen et al 2009).

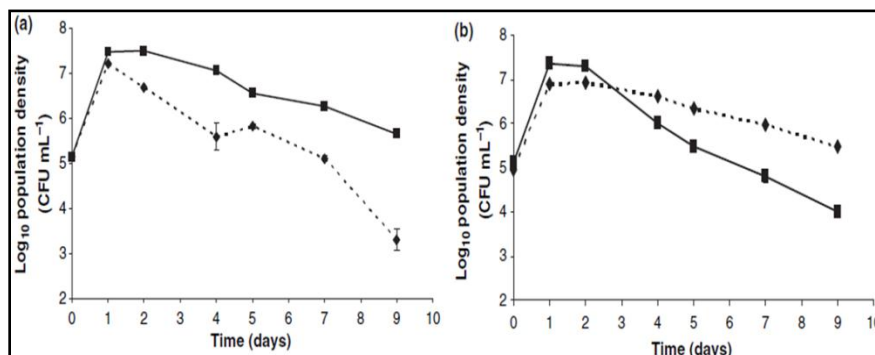


Figure 31 : Croissance de clones S et L (d'après Rozen et al 2009). Les deux types de clones ont été mis en culture seuls (a) ou ensemble (b). La croissance du clone L est représentée par des carrés et des lignes pleines; celle du clone S par des losanges et des lignes pointillées.

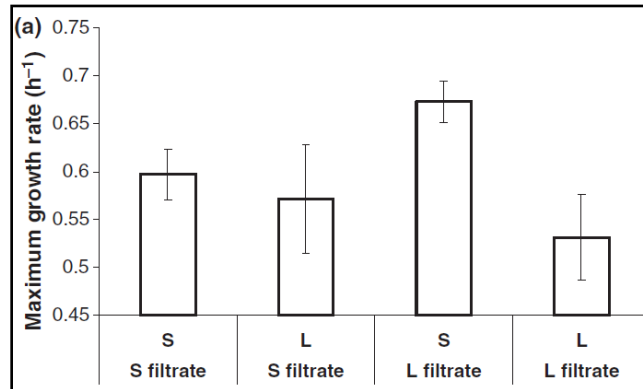


Figure 32 : Croissance des S et des L dans des filtrats de cultures de clones S et L (d'après Rozen et al 2009).

Le taux de croissance maximum des clones S et L a été mesuré dans un milieu qui consiste en un filtrat de culture en phase stationnaire de chacun des deux types de clones.

Un tel événement de divergence, suivi d'une co-existence à très long terme, ne s'est produit qu'une seule fois au sein des douze populations. L'émergence de la lignée S et la complexité des interactions entre les lignées S et L pourraient ainsi être associées à un phénomène de contingence historique et donc à une succession de multiples événements mutationnels. Ceux-ci auraient alors permis l'émergence et l'exploitation d'une nouvelle opportunité écologique liée à une construction de niche et à la saisonnalité de l'environnement.

Avant les travaux décrits dans cette thèse, aucune information n'était disponible ni sur les mécanismes permettant le maintien au long terme des deux lignées S et L, ni sur les bases génétiques de leur émergence. L'étude de ces mécanismes est d'autant plus importante qu'il s'agit d'expliquer l'émergence d'un événement unique de

polymorphisme, le plus stable jamais mis en évidence dans des conditions de laboratoire. L'objectif de ce travail de thèse vise ainsi à identifier les mécanismes de l'émergence et du maintien de ce polymorphisme. Les résultats seront présentés en trois chapitres :

- Le premier décrira ma participation au travail de post-doctorat de Mickaël Le Gac, au cours duquel nous nous sommes intéressés aux mécanismes assurant le maintien au long terme des lignées S et L. Cette partie a fait l'objet d'un article publié en 2012 dans la revue *Proceedings of the National Academy of Sciences USA*, et sera donc présentée sous forme de résumé avec l'article publié inséré.
- Le second chapitre présentera l'identification et l'analyse des mutations à la base de l'émergence de ce polymorphisme et plus particulièrement de la lignée S. Ces changements génétiques ont été corrélés aux phénotypes connus de la lignée S dans la population Ara-2 et ont été analysés pour essayer de comprendre en quoi cette diversification adaptative est spécifique de cette population. Cette partie est également présentée sous la forme d'une publication qui sera soumise prochainement, et contient aussi un résumé de ces résultats ainsi que des détails expérimentaux supplémentaires.
- Le troisième chapitre décrira l'analyse moléculaire d'une des mutations impliquées dans l'émergence de la lignée S. Cette mutation affecte un gène codant un régulateur global de la transcription chez *E. coli*, ici un régulateur de réponse d'un système à deux composants.

RESULTATS

RESULTATS PARTIE I

MAINTENANCE AU LONG TERME

D'UN EVENEMENT DE

DIVERSIFICATION BACTERIENNE

Au cours de ma thèse, j'ai participé au projet post-doctoral de Mickaël Le Gac, consistant en l'étude de la dynamique de co-existence des deux lignées S et L après leur émergence dans la population Ara-2 de l'expérience d'évolution au long terme (voir Introduction chapitre III). Cette étude a fait l'objet de la publication ci-jointe intitulée «Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*» (Mickaël Le Gac, Jessica Pluain, Thomas Hindré, Richard E. Lenski and Dominique Schneider, Proc. Natl Acad. Sci. USA, 2012, 109 : 9487-9492).

L'ensemble des résultats connus jusqu'alors sur la population Ara-2 ciblaient l'étude des facteurs écologiques de l'interaction entre clones des lignées S et L isolés à 18 000 générations (Rozen et al 2000, 2005, 2009). Cependant, les détails de la co-existence des deux écotypes, qui est la plus longue connue à ce jour dans des expériences d'évolution en laboratoire, restaient méconnus. L'enjeu est ici de comprendre comment ce polymorphisme, après son émergence, se maintient. En effet, à une extrême, il est possible qu'après la divergence écologique initiale, les deux écotypes aient continué à diverger pour occuper des niches écologiques différentes, conduisant à un équilibre stable entre eux. A l'autre extrême, des étapes successives d'évolution ont pu se produire conduisant plutôt à des interactions dynamiques au cours du temps.

Pour discriminer entre ces différentes possibilités, quatre clones de chacune des lignées S et L ont été isolés à chacun des trois temps suivants : 6500, 17 000 et 40 000 générations. Chaque série de quatre clones de la même lignée et du même temps a été

mélangée. Ces mélanges de clones évolués ont été comparés entre eux et par rapport au clone ancêtre pour de multiples phénotypes : la transcription de l'ensemble de leurs gènes par la détermination des profils globaux de transcription dans le milieu de culture correspondant aux conditions de l'expérience d'évolution, et leurs capacités de croissance dans plus de 50 environnements différents. Ces expériences ont ainsi permis d'analyser l'évolution phénotypique des deux lignées S et L après leur divergence par rapport à leur ancêtre commun. Des expériences de compétition entre les mêmes mélanges de clones S et L, isolés à un même temps évolutif ou à des temps différents, ont également été réalisées pour analyser si les interactions entre eux avaient évolué au cours du temps. Chacune de ces compétitions a été réalisée dans des conditions d'invasion réciproque, au cours desquelles chaque compétiteur (S ou L) est présent d'une part en minorité et d'autre part en majorité. Ainsi, la capacité d'invasion des S et des L a pu être testée non seulement entre des clones contemporains, mais aussi de temps antérieurs ou postérieurs.

Les profils globaux de transcription et les capacités de croissance montrent que les lignées S et L divergent fortement d'une part par rapport à leur ancêtre commun et d'autre part l'une par rapport à l'autre dès leur émergence à 6500 générations et ensuite au cours du temps (Figures 1 et 2 de l'article). Les données phénotypiques multiples obtenues suite à ces expériences ont été analysées plus finement en focalisant uniquement sur les gènes et les environnements révélant des différences entre les lignées S et L, et en regroupant ceux dont les profils d'expression et de croissance étaient similaires. Ainsi, un certain nombre de gènes ont révélé une expression spécifique soit chez les clones S soit chez les clones L, ce qui a permis d'identifier des

fonctions importantes pour chacune des lignées. Par exemple, la lignée S est caractérisée par une forte expression des gènes impliqués dans des voies métaboliques bien spécifiques comme le shunt du glyoxylate ou la voie d'Entner-Doudoroff. D'autre part, certains environnements révèlent des changements de capacités de croissance successivement chez les clones S puis chez les clones L au cours du temps, ce qui suggère que les clones L pourraient petit-à-petit empiéter sur les capacités cataboliques et donc les niches écologiques des S.

Cette dernière hypothèse a été corroborée par les résultats des expériences d'invasions réciproques (Figure 3 de l'article). Les clones de chacune des lignées S et L sont capables d'envahir, lorsqu'ils sont rares, les clones contemporains (isolés au même temps évolutif) de l'autre lignée, ce qui confirme les interactions négatives dépendantes de la fréquence entre les deux lignées, mais ce qui démontre également que chaque lignée occupe des niches écologiques suffisamment distinctes pour permettre leur co-existence. En revanche, une asymétrie a été observée dans les résultats des expériences d'invasions réciproques lorsque les compétiteurs S et L proviennent de générations différentes (Figure 3 de l'article). En effet, alors que les clones L sont capables d'envahir, lorsqu'ils sont rares, des clones S provenant à la fois de temps évolutifs antérieurs et postérieurs, les clones S ne sont capables d'envahir que des clones L antérieurs. Ces résultats des expériences d'invasion réciproque démontrent deux propriétés cruciales dans le maintien du polymorphisme entre les lignées S et L : d'une part, les clones L s'accrochent et envahissent les niches écologiques des clones S et d'autre part, les clones S sont capables de répondre et d'évoluer en conséquence en modifiant à nouveau leurs propres niches écologiques, bien que la nature de ces

différentes niches ne soit pas identifiée. Dans le cas contraire, cet accrochement de la lignée L aurait conduit à l'extinction de la lignée S.

Le maintien au long terme des deux lignées S et L se caractérise donc par un équilibre hautement dynamique de leurs interactions au cours du temps. Ainsi, après leur divergence liée à un événement de construction de niche (Rozen et al 2009), les deux lignées ont continué à évoluer. L'une d'elle s'accroche aux niches écologiques de l'autre, qui réagit de telle manière à éviter son extinction.

Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*

Mickaël Le Gac^{a,b}, Jessica Pluain^{a,b}, Thomas Hindré^{a,b}, Richard E. Lenski^{c,d,1}, and Dominique Schneider^{a,b,1}

^aLaboratoire Adaptation et Pathogénie des Microorganismes, Institut Jean Roget, Université Joseph Fourier, F-38041 Grenoble, France; ^bCentre National de la Recherche Scientifique, Unité Mixte de Recherche 5163, F-38041 Grenoble, France; and ^cDepartment of Microbiology and Molecular Genetics, and ^dBEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, MI 48824

Contributed by Richard E. Lenski, April 27, 2012 (sent for review January 17, 2012)

Closely related organisms usually occupy similar ecological niches, leading to intense competition and even extinction. Such competition also can promote rapid phenotypic evolution and ecological divergence. This process may end with the stable occupation of distinct niches or, alternatively, may entail repeated bouts of evolution. Here we examine two *Escherichia coli* lineages, called L and S, that coexisted for more than 30,000 generations after diverging from a common ancestor. Both lineages underwent sustained phenotypic evolution based on global transcription and resource utilization profiles, with L seeming to encroach over time on the catabolic profile of S. Reciprocal invasion experiments with L and S clones from the same or different generations revealed evolutionary changes in their interaction, including an asymmetry that confirmed the encroachment by L on the niche of the S lineage. In general, L and S clones from the same generation showed negative frequency-dependent effects, consistent with stable coexistence. However, L clones could invade S clones from both earlier and later generations, whereas S clones could invade only L clones from earlier generations. In this system, the long-term coexistence of competing lineages evidently depended on successive rounds of evolution, rather than on initial divergence followed by a static equilibrium.

experimental evolution | frequency-dependent selection | gene expression

The competitive exclusion principle precludes the stable coexistence of organisms occupying identical ecological niches (1–4). Closely related organisms tend to retain ancestral traits and thus occupy similar niches (5–9), and so competition is especially intense when they live in sympatry. Natural selection may then reduce their competition by driving apart those traits that determine their ecological niches, thereby allowing their stable coexistence through negative frequency-dependent selection. Examples of such ecological character displacement have been documented in many different organisms (10–14). However, the long-term evolutionary dynamics of competing lineages following such divergence are poorly understood (15). A theoretical model constructed to analyze the coevolution of competitors suggested that this process often led to extinctions (16), although a later study questioned some of the model's assumptions (17). In an empirical study, the interaction between two bacterial species was observed to evolve during a short-term experiment, leading to increased productivity of the community (18); however, that study examined neither the initial divergence of the bacteria nor the long-term fate of their interaction.

Consider the simple case of two recently diverged lineages that occupy slightly different ecological niches and are able to coexist stably owing to negative frequency-dependent effects. Their subsequent evolution could lead to several distinct scenarios. First, fitness might improve relative to earlier generations within, but not between, the lineages, sustaining their ecological differentiation. Second, competition between the lineages might be further reduced by character displacement, thereby promoting

their ecological divergence and stabilizing their interaction. Third, fitness may improve relative to previous generations both within and between lineages, leading to more complex changes in their interaction. For example, one lineage might encroach on the ecological niche of the other, and that encroachment in turn might cause the extinction of the affected lineage, or might lead to further evolution of the affected lineage that enables its persistence.

Here we studied the evolutionary dynamics of two bacterial lineages that diverged from their common ancestor and then coexisted for tens of thousands of generations in a long-term experiment. In that experiment, 12 populations of *Escherichia coli* were founded from the same strain and propagated in a glucose-limited minimal medium for more than 40,000 generations (19, 20). In one population, designated Ara-2, two lineages, named L and S, had diverged from their ancestor and from one another by 6,500 generations, and they coexisted thereafter (21–23). Their divergence and coexistence involved niche construction, in which organisms modify their environment in ways that promote their own or others' success (24–27). In particular, the two lineages compete for the limiting glucose, with the L lineage able to grow faster on the glucose but secreting a metabolic byproduct that S is better able to exploit (21, 23).

To characterize and better understand the evolutionary dynamics, we sampled four L and S clones from each of three time points: 6,500, 17,000, and 40,000 generations. We measured many phenotypes by quantifying and comparing their global transcription profiles in the medium where they evolved and their growth abilities in 51 different environments. We also performed reciprocal invasion experiments using the L and S populations from the same or different generations to assess how their ecological interaction had changed and what those changes might indicate about the evolutionary dynamics that allowed their long-term coexistence.

Results and Discussion

Expression Profiles and Growth Abilities of L and S Lineages. During the initial period through 6,500 generations, the bacteria underwent dramatic changes in global expression profiles relative to their ancestor (Fig. 14). More than 2,000 genes were differentially expressed between the ancestor and evolved bacteria

Author contributions: M.L.G., J.P., T.H., R.E.L., and D.S. designed research; M.L.G., J.P., T.H., and R.E.L. performed research; M.L.G. and D.S. analyzed data; and M.L.G., R.E.L., and D.S. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The data supporting this paper have been deposited in the DRYAD database, <http://datadryad.org> (accession: <http://dx.doi.org/10.5061/dryad.d7k5r5d2>). The microarray data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, <http://www.ncbi.nlm.nih.gov/geo> (accession no. GSE30639).

¹To whom correspondence may be addressed. E-mail: lenski@msu.edu or dominique.schneider@ujf-grenoble.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1207091109/-DCSupplemental.

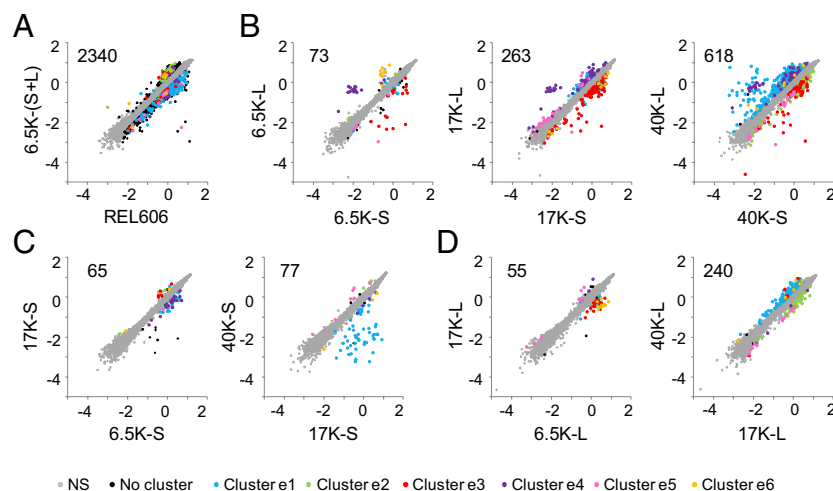


Fig. 1. Evolutionary dynamics of global transcription profiles for L and S lineages during 40,000 generations. (A) Comparison of the ancestral strain (REL606) with the combined S and L samples at 6,500 generations [6.5K – (S + L)]. (B) Comparisons of S and L samples at 6,500, 17,000, and 40,000 generations. (C) Comparisons of S samples from different generations. (D) Comparisons of L samples from different generations. Each point corresponds to a gene, and the values are \log_{10} -transformed expression levels. The number of genes with significantly different expression is shown for each comparison near the upper left corner. Gray symbols indicate genes without significant differences in expression in any of the eight comparisons. Those genes with significant differences in expression in at least one of the seven comparisons between two evolved samples were clustered according to their expression patterns, and the colored symbols indicate genes in expression clusters e1–e6 (Table S1). Black symbols indicate genes that either did not cluster or showed significant differences in expression only between the ancestor and evolved samples.

(including both L and S clones) according to the criteria specified in *Materials and Methods*. These changes reflect substantial reorganization of gene regulatory networks, and such global changes were observed repeatedly in this long-term experiment (28–30). The growth abilities of the bacteria also changed over this same period in 16 of the 51 environments that yielded informative data (Fig. 2A).

We also compared the expression profiles and growth abilities of L and S clones sampled at identical time points. After 6,500 generations, the L and S clones showed significant differences in expression for only 73 genes (Fig. 1B), although the magnitude of the differences (deviations from the 45-degree line) tended to be large compared with their differences relative to the ancestor.

The number of genes differentially expressed between the L and S lineages further increased to 263 after 17,000 generations and to 618 after 40,000 generations. The number of test environments in which the growth of the L and S lineages differed showed no clear trend, with 20 differences at 6,500 generations, 27 at 17,000 generations, and 18 at 40,000 generations (Fig. 2B). However, the differences between the two lineages tended to be greater in the later generations, and the environments in which they differed also changed, as described below.

Cluster Analyses. The differences in gene expression and growth yields demonstrate that many physiological properties of the L and S lineages evolved during the thousands of generations in

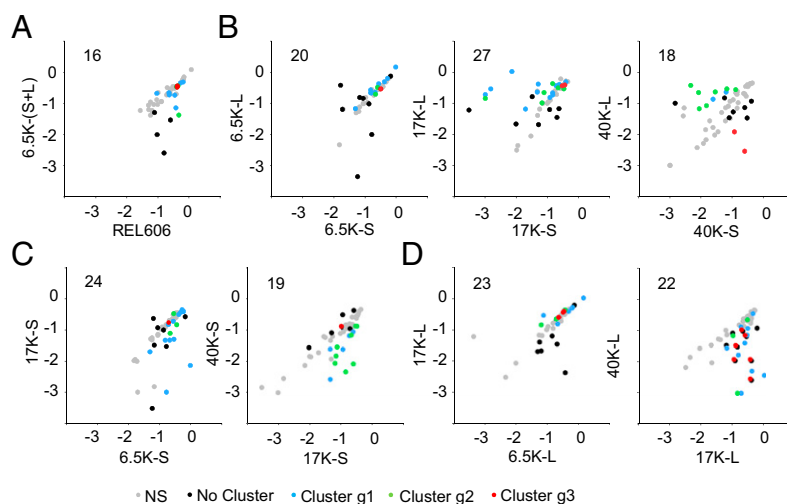


Fig. 2. Evolutionary dynamics of growth abilities across 51 environments for L and S lineages during 40,000 generations, as described in Fig. 1. Each point corresponds to a different environment, and the values are \log_{10} -transformed growth yields based on optical densities. The number of environments with significant differences in growth is shown for each comparison near the upper left corner. Gray symbols indicate environments without significant differences in growth in any of the eight comparisons. Those environments demonstrating significant differences in growth in at least one comparison between two evolved samples were clustered based on their growth patterns, and the different colored symbols indicate environments in growth clusters g1–g3 (Table S2). Black symbols indicate environments that either did not cluster or showed significant differences only between the ancestor and evolved samples.

which they coexisted. We performed cluster analyses of the gene expression and growth datasets to explore the evolution of these multidimensional phenotypes, with a focus on understanding the differences between the L and S lineages and how they evolved over time. Thus, we analyzed those genes and environments (797 and 48, respectively) for which we observed a significant difference in expression or growth in at least one of the seven comparisons between evolved clones, excluding genes and environments that showed no significant difference in any comparison, as well as those for which the only difference was between the ancestral and evolved bacteria. These analyses identified six clusters of genes with differential expression patterns (e1–e6) and three clusters of environments that produced differential growth patterns (g1–g3).

Expression cluster e1 includes 367 genes of diverse functions (Table S1) for which mRNA levels were significantly higher in the L lineage than in the S lineage at generation 40,000 (Fig. 1B). Some of these genes showed substantially reduced expression in the S lineage between 17,000 and 40,000 generations (Fig. 1C), but most exhibited slightly increased expression in the L lineage over the same period (Fig. 1D). Many of the genes with strongly reduced expression in S were found in two chromosomal regions that had been deleted by mutations in that lineage (Table 1). The detection of apparent low expression of the deleted genes indicates background noise (including any cross-hybridization) in the microarrays used to quantify gene expression. Cluster e2 has 133 genes, also with diverse cellular functions, that were expressed at lower levels in the L lineage than in the S lineage at generation 40,000 (Fig. 1B), largely as a consequence of changes in L between 17,000 and 40,000 generations (Fig. 1D).

Expression clusters e5 and e6 together include 84 genes that showed more transient changes. The genes in e5 were expressed at slightly higher levels in the L lineage than in the S lineage after 17,000 generations, whereas the reverse pattern was seen at 40,000 generations (Fig. 1B). The genes in cluster e6 had higher expression in L than in S at 6,500 generations, but this pattern was largely eliminated by 40,000 generations (Fig. 1B), primarily as a consequence of reduced expression of these genes in the L lineage between 6,500 and 17,000 generations (Fig. 1D).

Expression clusters e3 and e4 together include another 176 genes for which the mRNA levels differed significantly between the L and S lineages (Fig. 1B). In contrast to the clusters discussed above, the expression of genes in these two clusters remained largely constant between generations 6,500 and 40,000 within each lineage (Fig. 1C and D). These fixed differences likely include or reflect those physiological traits important for the early ecological specialization of each lineage. Many of the genes that were more highly expressed in the S lineage (cluster e3) are involved in the Entner–Doudoroff pathway (31), an alternative to glycolysis that also catabolizes glucose to pyruvate, and in the glyoxylate cycle (32), a bypass of the tricarboxylic acid cycle that enables growth on two-carbon compounds. These differences may contribute to the lower growth rate of the S cells on glucose and to the ability of the S cells to exploit metabolic byproducts secreted by L cells (21, 23). Some of the genes that were more highly expressed in the L lineage (cluster e4) are involved in glycerol metabolism; these include *manXYZ*, which encodes a sugar transporter controlled by the glucose-sensing

global regulator Mlc (33). The ManXYZ protein has been identified as a secondary glucose transporter (33), and its increased expression may contribute to the faster growth of the L cells on glucose.

Turning to the cluster analyses of differences in growth across environments (Table S2), clusters g2 and g3 together include 18 environments in which the measured yield of the L lineage increased or decreased relative to the S lineage (Fig. 2B). Most of these differences resulted from declines in the yield of one lineage, as opposed to higher yields in the other lineage (Fig. 2C and D). The finding that many differences reflect diminished growth in the test environments may indicate trade-offs caused by pleiotropic effects of beneficial mutations, accumulation of mutations in genes under relaxed selection in the long-term experiment, or both (34). This population evolved a mutator phenotype even before the L and S lineages diverged (22, 35), and the higher mutation rate in mutators allows for faster accumulation of mutations in those genes that affect nonessential functions (34, 36).

The growth cluster g1 includes 18 other environments in which the growth abilities of the two lineages decreased sequentially, first declining in S between generations 6,500 and 17,000 (Fig. 2C) and then declining in L between generations 17,000 and 40,000 (Fig. 2D). None of these test environments directly reflects the conditions of the medium and metabolic byproducts that were present during the evolution experiment. However, this cluster indicates that some traits that early on were specific to the S lineage later evolved in the L lineage, raising the intriguing possibility that L was encroaching on the ecological niche of S.

Reciprocal Invasion Experiments. To test the encroachment hypothesis, we ran competition experiments using the same sets of L and S clones. Experiments were performed using a reciprocal-invasion design—in which the initial ratio of the two lineages was either 1:9 or 9:1, to determine whether they could stably coexist or, alternatively, whether one would exclude the other (21, 37).

When the competitors were from the same generation, each lineage had a significant advantage when it was initially rare (Fig. 3, side-facing black wedges), which demonstrates a negative frequency-dependent interaction (38). That relationship also confirms that the lineages occupied sufficiently distinct ecological niches to allow them to coexist. When the competitors came from different generations, however, we observed a striking asymmetry in outcomes. In particular, both L and S clones could invade the alternative lineage from the earlier generation when rare (Fig. 3, downward-angled red and blue wedges). Moreover, initially rare L were able to invade S from later generations (Fig. 3, upward-angled blue wedges), but S could not invade L from later generations (Fig. 3, upward-angled dashed red lines); that is, earlier S clones were significantly less fit than later L clones, even when the S clones started at low frequency.

This last result demonstrates that the S lineage would have been driven to extinction by the continued evolution of the L lineage had S itself not evolved in response. This result also implies that L encroached on the ecological niche of S, because otherwise S would not have been affected by the further

Table 1. Large genomic deletions in S lineage at 40,000 generations

Deletion	Size, kbp	Coordinates*	Deleted genes
1 ($\Delta kpsD\text{--}glcB$)	~35	3015256–3050311	<i>kpsD</i> , <i>kpsU</i> , <i>kpsC</i> , <i>kpsS</i> , <i>insB</i> , <i>kpsT</i> , <i>kpsM</i> , <i>yghD</i> , <i>yghE</i> , <i>epsF</i> , <i>epsE</i> , <i>epsD</i> , <i>yghF</i> , <i>yghG</i> , <i>pppA</i> , <i>yghI</i> , <i>yghK</i> , <i>glcB</i>
2 ($\Delta insB\text{--}ybdK$)	~41	547701–588493	<i>insB</i> , <i>insA</i> , <i>essD</i> , <i>ybcS</i> , <i>rzpD</i> , <i>borD</i> , <i>ynfO</i> , <i>nohA</i> , <i>appY</i> , <i>ompT</i> , <i>envY</i> , <i>ybcH</i> , <i>nfrA</i> , <i>nfrB</i> , <i>yhhI</i> , <i>cusS</i> , <i>cusR</i> , <i>cusC</i> , <i>cusF</i> , <i>cusB</i> , <i>cusA</i> , <i>pheP</i> , <i>ybdG</i> , <i>nfnB</i> , <i>ybdF</i> , <i>ybdI</i> , <i>ybdK</i>

*Coordinates are based on the ancestral strain's genome sequence (39).

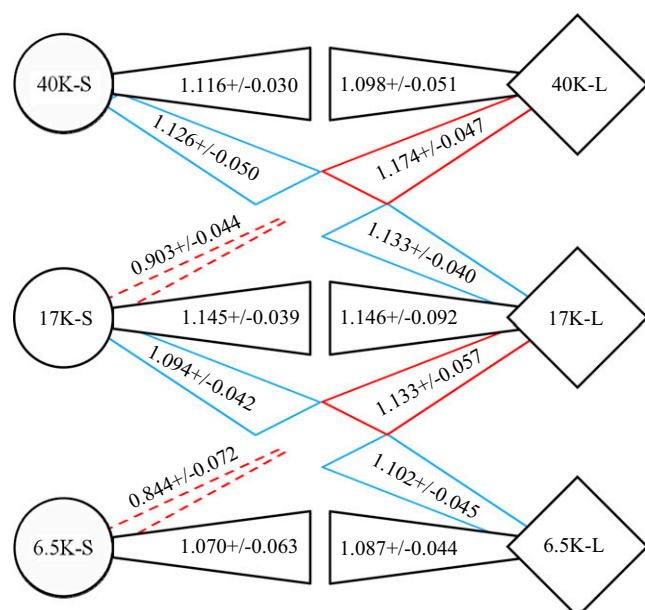


Fig. 3. Changes in ecological interactions between L and S lineages during 40,000 generations. Sets of clones from the L (diamonds) and S (circles) lineages were sampled at 6,500, 17,000, and 40,000 generations. Each set from each time point competed against two or three other sets from the same or different generations. Black, red, and blue wedges indicate competitions between contemporary lineages, between later L and earlier S, and between earlier L and later S, respectively. Competitions were performed at two starting ratios, with each competitor either initially rare (10%) or common (90%). Wedges that expand (solid lines) or narrow (dashed lines) from each competitor indicate its advantage (fitness > 1) or disadvantage (fitness < 1) when initially rare. Values represent the mean fitness (with 95% confidence interval) of the rare competitor relative to the common competitor. With 15-fold replication of each competition, all values differ significantly from unity.

adaptation of L. This encroachment could, in principle, result from the improved performance of L on some shared resource (including glucose), from a newly evolved capacity of L to use a secreted metabolite that was previously used only by S, or from some change in an allelopathic effect whereby the organisms produce substances that harm their competitors.

Synthesis and Perspective. The coexistence of two competitors may be dynamically stable in a strictly ecological context, but evolutionary changes in one or both populations may affect the stability of the interaction over longer periods. Evolution might enhance stability by promoting character displacement, thereby reducing the intensity of competition between the lineages, or evolution might destabilize the interaction if one lineage encroaches on the other's niche. In the latter case, the encroached-upon lineage might be driven to extinction, or it might evolve in some way that allows it to persist. Although previous studies have documented character displacement and the resulting coexistence of closely related lineages (10–14), the long-term dynamics of these interactions have not received much attention.

Our ability to analyze these dynamics benefitted from two features of the system under study. First, owing to the short generation times of bacteria, we could observe both the initial divergence of lineages (over several thousand generations) and their subsequent evolution (for tens of thousands of generations). Second, bacteria can be frozen and revived, which allowed us to compete organisms that lived at different times and thereby observe the asymmetry in their evolving interaction. Our results illustrate the rich and complex evolutionary dynamics that can occur between two recently diverged lineages even in a simple

environment. The initial diversification involved the process of niche construction (24–27, 37) in which one lineage produced a byproduct that the other lineage was better able to exploit as a resource, allowing the two lineages to occupy distinct niches and thereby coexist (21–23). The two lineages then continued to evolve, with both undergoing substantial changes in gene expression profiles and growth abilities under various conditions. Their competitive interaction also evolved as one lineage encroached on the ecological niche of the other lineage, which in turn responded in ways that allowed it to persist. More generally, the character displacement that enables closely related organisms to coexist need not be an evolutionary endpoint but instead may be only the first step in a long evolutionary dance.

Materials and Methods

Strains. All strains were sampled from population Ara-2 of the *E. coli* long-term evolution experiment (19, 21–23) or were marked variants thereof. In this population, two lineages, L and S, had diverged from one another by generation 6,500 and then coexisted until at least generation 40,000. Based on derived mutations specific to each lineage (22), we sampled four clones from each lineage from frozen samples obtained at three time points during the evolution experiment: generations 6,500 (6.5K), 17,000 (17K), and 40,000 (40K). The four clones from each lineage and generation were mixed, generating the six evolved samples used in this study, designated 6.5K-S, 6.5K-L, 17K-S, 17K-L, 40K-S, and 40K-L. These samples were used for global transcription profiling, growth assays, and competition experiments.

The Ara-2 population was founded by an ancestral strain, REL606, that cannot use arabinose as a carbon source (19, 39), and all evolved clones were also phenotypically Ara[−]. To distinguish competitors during the competition experiments, we isolated spontaneous Ara⁺ mutants of each of the 24 evolved clones, as described elsewhere (21). Again, the four Ara⁺ clones from each lineage and generation were mixed before the competition experiments. The arabinose utilization marker is effectively neutral under the conditions of the long-term evolution experiment and in the L and S genetic backgrounds (19, 21).

Global Transcription Profiles. The seven samples, including the six evolved mixtures and the ancestral strain, were grown overnight in 2 mL of Davis minimal (DM) medium (19) containing 1 mg/mL of glucose (DM1000), then diluted 10⁴ fold into 10 mL of DM medium containing 25 μg/mL of glucose (DM25) in 50-mL flasks and incubated for 24 h at 37 °C while shaking at 120 rpm. To obtain sufficient amounts of RNA, these cultures were diluted 1:100 into 200 mL of DM25 in 1-L flasks and incubated until they reached mid-exponential phase (6 h for the ancestral strain and 4.5 h for the evolved mixtures). Pellets of bacterial cells were produced and total RNAs extracted using the Qiagen RNeasy Kit, following the manufacturer's protocol. We grew six replicate cultures and analyzed six replicate RNA extractions for each sample. Biotinylated cDNA samples were prepared according to the standard Affymetrix GeneChip protocol for bacteria, and the samples were then hybridized to the Affymetrix *E. coli* Genome 2.0 Microarray for 16 h at 45 °C by Beckman Coulter Genomics. Quality checks and normalization were performed using the Affymetrix GeneChip Command Console and Expression Console according to standard Affymetrix protocols. In brief, the intensity data were subjected to global scaling to compare data from multiple probe arrays.

Identifying Differentially Expressed Genes. After removing the intergenic regions from the dataset, we identified differentially expressed genes using the optimal discovery procedure (40) implemented in the EDGE v1.1.208 software (41). We performed eight comparisons, including the ancestor vs. the combined S and L samples from generation 6,500 [6.5K – (S + L)], three between-lineage comparisons at different generations (6.5K-S vs. 6.5K-L, 17K-S vs. 17K-L, and 40K-S vs. 40K-L); and four within-lineage comparisons across generations (6.5K-S vs. 17K-S, 17K-S vs. 40K-S, 6.5K-L vs. 17K-L, and 17K-L vs. 40K-L). Each analysis was performed with 200 permutations. Genes were scored as differentially expressed when the false discovery rate (FDR) (42) had a *q* value < 0.05 and the magnitude of the change was greater than twofold. With multiple testing, the FDR estimates the probability, *q*, of one gene being a false-positive, given the overall distribution of the ordinary *P* values obtained for genes analyzed one at a time. Those genes identified as differentially expressed in at least one of the seven comparisons between two evolved samples (797 genes in total) were clustered according to their expression patterns (Fig. 1) using the QT CLUST algorithm (43) implemented

with TM4 software (44). We used the Pearson correlation metric distance with a maximum cluster diameter of 0.65 and a minimum cluster size of 25; similar clusters were obtained when we varied the parameters. Genes were classified by Gene Ontology (GO) cellular process categories (45) and also by regulons according to RegulonDB (46). We tested the overrepresentation of GO categories and regulons (Table S1) using one-tailed Fisher's exact tests that compared the number of differentially expressed genes in each of the six clusters with the total number of differentially expressed genes. These tests were performed only when two or more genes within a GO category or regulon were differentially expressed in the focal comparison. The tests were performed multiple times; thus, FDRs were estimated using QVALUE software (42), and categories were identified as overrepresented only for q values <0.05 .

Characterization of Deleted Chromosomal Regions in the S Sineage at 40,000 Generations.

Two sets of physically adjacent genes exhibited large reductions in expression in the S lineage at 40,000 generations (Table 1). These observations suggested that the genes might have been deleted from the genome, with the residual expression reflecting background noise (including any cross-hybridization) in the microarrays. We performed PCR experiments that confirmed the deletion of two large regions, and we identified the precise endpoints of each deletion. We designed primer pairs complementary to the upstream and downstream sequences of each region, which were then used with genomic DNA from a 40,000-generation S clone. For deletion 1 ($\Delta kpsD\text{-}glcB$), the pairs were 5'-GCGAAATAAGCGAAAACGAG-3' and 5'-GTTGCAAAATGGCGATACCT-3'. For deletion 2 ($\Delta insB\text{-}ybdK$), the pairs were 5'-AGTCTCCGACCAGAAGCGTA-3' and 5'-TTACCATAGGCGACCTGACC-3'.

Growth Assays Using Biolog Microplates. The same seven biological samples were grown overnight in 2 mL of DM1000, then diluted 10^4 -fold into 10 mL of DM25 in 50-mL flasks and incubated for 24 h at 37 °C with shaking at 120 rpm. For each sample, 2 μ L of culture and 100 μ L of sterile DM medium were added to each well of a Biolog GENIII microplate. The plates were then incubated at 37 °C in a microplate reader (Infinite M200; Tecan), and the optical density at 600 nm (OD_{600}) was measured periodically for each well. The wells contained 71 carbon sources and 23 other chemicals for assessing substrate use and chemical sensitivity, respectively. Each assay was performed in triplicate. In most cases, we quantified growth yield as the maximum OD_{600} reached within the first 10 h; however, under four conditions (fusidic acid, D-saccharic acid, L-malic acid, and sodium butyrate), growth began only after a long lag phase, and for these we quantified growth yield

as the OD_{600} reached after 24 h. We excluded from the analysis all conditions for which OD_{600} did not reach 0.06 in any sample, leaving a total of 51 conditions that were informative.

To analyze the growth data, we performed the same eight comparisons as on the gene expression data. Differences in growth abilities were tested using Wilcoxon's rank-sum test. Those environments that showed significant growth differences in at least one of the seven comparisons between two evolved samples (48 environments in total) were clustered based on their patterns of growth using the QT CLUST algorithm (43) in TM4 software (44). We used the Pearson correlation metric distance with a maximum cluster diameter of 0.40 and a minimum cluster size of five; similar clusters were obtained when these parameters were varied.

Competition Experiments. We performed seven sets of pairwise competition experiments to determine whether the L and S lineages from the same or different generations could invade one another when rare. The pairs included 6.5K-S vs. 6.5K-L, 17K-S vs. 17K-L, 40K-S vs. 40K-L, 6.5K-S vs. 17K-L, 6.5K-L vs. 17K-S, 17K-S vs. 40K-L, and 17K-L vs. 40K-S. Each competition experiment ran for 24 h (one complete serial transfer cycle) in the same medium and conditions as used in the long-term evolution experiment (19). The competitors were mixed at two initial ratios (1:9 and 9:1), and 15 replicate experiments were conducted for each pair and each initial ratio. The L and S competitors were distinguished on the basis of an arabinose utilization marker that is neutral (i.e., has no significant effect on fitness) under these conditions (19, 21), with each marker state used for each competitor in approximately half of the replicates. This marker allows the scoring of strains by colony color when plated on an appropriate medium. The fitness of one competitor relative to the other was calculated as the ratio of their net growth rates during the competition experiment (19, 21). The Student t test was used to evaluate whether the mean fitness differed from the null hypothetical expectation of 1 for each pair of competitors at each initial ratio.

ACKNOWLEDGMENTS. We thank D. Schluter and P. B. Rainey for helpful comments on this paper. This work was supported by Agence Nationale de la Recherche Program "Blanc" ANR-08-BLAN-0283-01, Centre National de la Recherche Scientifique, and Université Joseph Fourier (D.S.); and by National Science Foundation DEB-1019989 and BEACON Center for the Study of Evolution in Action National Science Foundation Cooperative Agreement DBI-0939454 (R.E.L.). M.L.G. was supported by a research fellowship from the Agence Nationale de la Recherche. J.P. was supported by a doctoral fellowship from the French Ministry of Research and the Université Joseph Fourier.

- Gause GF (1932) Experimental studies on the struggle for existence. I: Mixed population of two species of yeast. *J Exp Biol* 9:389–402.
- Hutchinson GE (1957) Concluding remarks. *Cold Spring Harb Symp Quant Biol* 22: 415–427.
- Hardin G (1960) The competitive exclusion principle. *Science* 131:1292–1297.
- Holt RD (2009) Bringing the Hutchinsonian niche into the 21st century: Ecological and evolutionary perspectives. *Proc Natl Acad Sci USA* 106(Suppl 2):19659–19665.
- Peterson AT, Soberon J, Sanchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. *Science* 285:1265–1267.
- Wiens JJ, Graham CH (2005) Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annu Rev Ecol Syst* 36:519–539.
- Losos JB (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecol Lett* 11:995–1003.
- Gómez JM, Verdú M, Perfectti F (2010) Ecological interactions are evolutionarily conserved across the entire tree of life. *Nature* 465:918–921.
- Wiens JJ, et al. (2010) Niche conservatism as an emerging principle in ecology and conservation biology. *Ecol Lett* 13:1310–1324.
- Brown WL, Wilson EO (1956) Character displacement. *Syst Zool* 5:49–64.
- Schluter D (2000) *The Ecology of Adaptive Radiations* (Oxford Univ Press, Oxford).
- Grant PR, Grant BR (2008) *How and Why Species Multiply* (Princeton Univ Press, Princeton).
- Pfennig KS, Pfennig DW (2009) Character displacement: Ecological and reproductive responses to a common evolutionary problem. *Q Rev Biol* 84:253–276.
- Brockhurst MA, Hochberg ME, Bell T, Buckling A (2006) Character displacement promotes cooperation in bacterial biofilms. *Curr Biol* 16:2030–2034.
- Ricklefs RE (2010) Evolutionary diversification, coevolution between populations and their antagonists, and the filling of niche space. *Proc Natl Acad Sci USA* 107:1265–1272.
- Rummel J, Roughgarden J (1985) A theory of faunal buildup for competition communities. *Evolution* 39:1009–1033.
- Taper ML, Case TJ (1992) Models of character displacement and the theoretical robustness of taxon cycles. *Evolution* 46:317–333.
- Hansen SK, Rainey PB, Haagenen JAJ, Molin S (2007) Evolution of species interactions in a biofilm community. *Nature* 445:533–536.
- Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I: Adaptation and divergence during 2,000 generations. *Am Nat* 138:1315–1341.
- Barrick JE, et al. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461:1243–1247.
- Rozen DE, Lenski RE (2000) Long-term experimental evolution in *Escherichia coli*, VIII: Dynamics of a balanced polymorphism. *Am Nat* 155:24–35.
- Rozen DE, Schneider D, Lenski RE (2005) Long-term experimental evolution in *Escherichia coli*, XIII: Phylogenetic history of a balanced polymorphism. *J Mol Evol* 61:171–180.
- Rozen DE, Philippe N, de Visser JA, Lenski RE, Schneider D (2009) Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecol Lett* 12:34–44.
- Odling-Smee FJ, Laland KN, Feldman MW (1996) Niche construction. *Am Nat* 147: 641–648.
- Laland KN, Odling-Smee FJ, Feldman MW (1999) Evolutionary consequences of niche construction and their implications for ecology. *Proc Natl Acad Sci USA* 96:10242–10247.
- Erwin DH (2008) Macroevolution of ecosystem engineering, niche construction and diversity. *Trends Ecol Evol* 23:304–310.
- Kylafis G, Loreau M (2008) Ecological and evolutionary consequences of niche construction for its agent. *Ecol Lett* 11:1072–1081.
- Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci USA* 100: 1072–1077.
- Crozat E, Philippe N, Lenski RE, Geiselmann J, Schneider D (2005) Long-term experimental evolution in *Escherichia coli*, XII: DNA topology as a key target of selection. *Genetics* 169:523–532.
- Philippe N, Crozat E, Lenski RE, Schneider D (2007) Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *Bioessays* 29:846–860.
- Peekhaus N, Conway T (1998) What's for dinner?: Entner–Doudoroff metabolism in *Escherichia coli*. *J Bacteriol* 180:3495–3502.
- Kornberg H (2000) Krebs and his trinity of cycles. *Nat Rev Mol Cell Biol* 1:225–228.
- Plumbridge J (2002) Regulation of gene expression in the PTS in *Escherichia coli*: The role and interactions of Mlc. *Curr Opin Microbiol* 5:187–193.
- Cooper VS, Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* 407:736–739.

35. Sniegowski PD, Gerrish PJ, Lenski RE (1997) Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387:703–705.
36. Funchain P, et al. (2000) The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* 154:959–970.
37. Rainey PB, Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* 394:69–72.
38. Snyder TP, Ayala FJ (1979) Frequency-dependent selection at the PGM-1 locus of *Drosophila pseudoobscura*. *Genetics* 92:995–1003.
39. Jeong H, et al. (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3). *J Mol Biol* 394:644–652.
40. Storey JD, Dai JY, Leek JT (2007) The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Biostatistics* 8:414–432.
41. Leek JT, Monsen EC, Dabney AR, Storey JD (2006) EDGE: Extraction and analysis of differential gene expression. *Bioinformatics* 22:507–508.
42. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 100:9440–9445.
43. Heyer LJ, Kruglyak S, Yooseph S (1999) Exploring expression data: Identification and analysis of coexpressed genes. *Genome Res* 9:1106–1115.
44. Saeed AI, et al. (2003) TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.
45. Riley M, et al. (2006) *Escherichia coli* K-12: A cooperatively developed annotation snapshot—2005. *Nucleic Acids Res* 34:1–9.
46. Gama-Castro S, et al. (2008) RegulonDB (version 6.0): Gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* 36(Database issue):D120–D124.

Supporting Information

Le Gac et al. 10.1073/pnas.1207091109

Table S1. Cluster analysis of genes differentially expressed between the L and S lineages within and across generations

		Cluster (no. of genes)					
		e1 (367)	e2 (133)	e3 (96)	e4 (80)	e5 (57)	e6 (27)
Cellular functions	Carbohydrate catabolism	Iron transport	Glyoxylate cycle	Glycerol metabolism	Enterobactin biosynthesis	Glutamate biosynthesis	
	Flagella biogenesis	Enterobactin biosynthesis	Nitrogen metabolism	Sulfur metabolism	Nonribosomal peptide biosynthesis	Aspartate biosynthesis	
	Amino sugar biosynthesis	RNA modification	Entner–Doudoroff pathway		Iron transport	Polysaccharide catabolism	
	Taxis	Nucleobase, nucleoside, and nucleotide interconversion	Glutamate biosynthesis		Nucleobase, nucleoside, and nucleotide interconversion		
	Fatty acid oxidation	Purine biosynthesis					
	Peptidoglycan metabolism	Response to temperature					
	Amino acid catabolism						
	Glycerol metabolism						
	Methylglyoxal metabolism						
Regulators	<i>crp, csfR, cusR, fadR, flhC, flhD, fucR, gutM, lrp, rpiR, srlR</i>	<i>fur, nadR, purR</i>	<i>fruR, glnG, gntR, iclR, nac, phoP, rpoN</i>	<i>cysB, mlc, glpR, nagC, nanR</i>	<i>fur</i>	<i>malT</i>	

Columns show cellular functions and regulators that are overrepresented in the six gene expression clusters e1–e6. The number of genes in each cluster is shown after the cluster label. In addition, 37 genes showed variable expression between evolved samples but did not cluster. Many other genes showed no significant differences between samples or significant differences only between the ancestral and evolved samples, and those genes were not included in the cluster analysis. Cellular functions correspond to the Gene Ontology cellular process categories (1); regulators were identified using RegulonDB (2).

1. Riley M, et al. (2006) *Escherichia coli* K-12: A cooperatively developed annotation snapshot—2005. *Nucleic Acids Res* 34:1–9.
2. Gama-Castro S, et al. (2008) RegulonDB (version 6.0): Gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* 36(Database issue):D120–D124.

RESULTATS PARTIE II

BASES GENETIQUES DE

L'EMERGENCE DE

L'EVENEMENT DE

DIVERSIFICATION BACTERIENNE

Le cœur de mes travaux de thèse, présenté dans ce chapitre (et dans le suivant), a consisté à comprendre les bases génétiques et moléculaires qui ont permis l'émergence du polymorphisme conduisant aux lignées S et L dans la population Ara-2. Ces travaux viennent ainsi compléter ceux décrits dans le chapitre précédent sur la co-existence au long terme des deux lignées. Ce chapitre est également présenté sous la forme d'une publication ci-jointe intitulée «Multi-step emergence of a long-term polymorphism in *Escherichia coli*» (Jessica Plucain, Mickaël Le Gac, Olivier Tenaillon, Stéphane Cruveiller, Claudine Médigue, Richard E. Lenski, Thomas Hindré & Dominique Schneider), que nous souhaitons prochainement soumettre à la revue Nature. Cette publication en préparation compile la plupart des résultats obtenus concernant l'émergence de la lignée S. Cependant, des résultats supplémentaires ont été obtenus et sont présentés à la fin de cette partie, après l'article.

Si les paramètres écologiques impliqués dans le maintien au long terme du polymorphisme ont jusqu'alors été appréhendés (Rozen et al 2000, 2005, 2009, Le Gac et al 2012), les bases génétiques de cet événement de diversification sont totalement inconnues. Pour les identifier, les génomes de clones isolés au cours de l'évolution de la population Ara-2 ont été séquencés et analysés. Les résultats présentés ici se focalisent sur les premières 6500 générations qui correspondent à la période pré-divergence et d'émergence des lignées S et L, afin de déterminer les événements mutationnels qui en sont responsables. En effet, cette expérience d'évolution au long terme offre, de par sa conception, une occasion unique de décortiquer finement les étapes conduisant à l'émergence de la diversité au sein d'une population homogène.

Sur la base de ces analyses de séquences génomiques, l'ensemble des mutations associées à la période d'émergence du polymorphisme ont été identifiées : 68 mutations sont spécifiques de la période précédant la divergence des deux lignées (période pré-divergence), et 55 et 36 mutations sont spécifiques des lignées S et L, respectivement (Figure 1 et Tableaux supplémentaires 2 à 4 de l'article). Ces mutations ont alors été analysées par rapport aux profils globaux de transcription caractéristiques des lignées S et L (Le Gac et al 2012, voir Résultats Partie I). Parmi les 68 mutations spécifiques de la période pré-divergence, nous nous sommes focalisés sur celles affectant des gènes codant des régulateurs globaux de l'expression des gènes, car les profils de transcription révèlent de très importants changements d'expression pendant cette période pré-divergence. De plus, les gènes de régulateurs globaux ont été démontrés comme étant les cibles majeures de la sélection naturelle dans ces populations (Philippe et al 2007, Hindré et al 2012). D'autre part, les profils globaux de transcription ont identifié des différences d'expression des gènes impliqués dans le shunt du glyoxylate et la voie d'Entner-Doudoroff spécifiquement au sein de la lignée S. Sur la base de la comparaison entre les données génomiques et transcriptomiques, trois mutations candidates ont été identifiées et analysées plus en détail. Elles affectent les trois gènes *spoT*, *arcA* et *gntR*, tous trois codant des régulateurs de la transcription impliqués dans le contrôle du métabolisme central chez *E. coli*. La mutation de *spoT* est la première des trois à être substituée dans la population Ara-2, au cours de la période pré-divergence, suivie de la mutation de *arcA* puis de celle de *gntR* qui sont présentes spécifiquement chez les clones de la lignée S. La protéine SpoT joue un rôle majeur dans la réponse stringente et la réponse aux carences nutritives qui contrôlent les transitions entre phase

exponentielle de croissance et phase stationnaire (voir Introduction Partie I section II.1). Ces transitions sont importantes dans la saisonnalité de l'expérience d'évolution, qui est elle-même cruciale pour les interactions entre S et L (Rozen et al 2009). La protéine ArcA est le régulateur de réponse du système à deux composants ArcA/ArcB impliqué notamment dans la régulation des gènes codant des enzymes du cycle de Krebs et du shunt du glyoxylate, en fonction des conditions d'oxygénation du milieu (voir Introduction Partie I section III). La protéine GntR est le répresseur des gènes de la voie de dégradation alternative du glucose appelée Entner-Doudoroff, voie normalement réprimée chez *E. coli* dans les conditions de laboratoire. Cette voie de dégradation semble être activée dans le gros intestin où des sources de carbone comme le gluconate, le glucuronate ou le galacturonate sont présentes en quantité abondante et sont dégradées via cette voie (Peekhaus et Conway 1998).

Nous avons alors construit des souches isogéniques, avec et sans ces mutations, dans les contextes génétiques ancestral et évolué de la lignée S. La restauration des allèles ancestraux de chacun des gènes *arcA* et *gntR* dans le contexte génétique évolué de la lignée S conduit à l'abolition des phénotypes caractéristiques de la lignée S (Figure 2 et Figure supplémentaire 1 de l'article). En revanche, l'introduction des allèles évolués des trois gènes *spoT*, *arcA* et *gntR* dans le contexte génétique ancestral confère à cet ancêtre des caractéristiques phénotypiques similaires à un clone S, à savoir des interactions négatives dépendantes de la fréquence et la capacité de se développer dans un surnageant de culture d'un clone L (Figures 2 et 3, et Figure supplémentaire 2 de l'article).

Les gènes *spoT*, *arcA-arcB* et *gntR* sont affectés par des mutations, correspondant à d'autres allèles, dans 7, 10 et 1 des autres populations de l'expérience d'évolution au long terme. Pourtant, seule la population Ara-2 révèle la présence de la divergence vers les lignées S et L. Nous avons ainsi voulu tester si l'originalité de l'événement de diversification qui est apparu dans cette population résidait dans la spécificité des allèles sélectionnés. Nous avons ainsi remplacé la mutation *spoT* apparue dans la population Ara-2 par une mutation *spoT* présente dans une autre population et affectant un autre acide aminé de la protéine SpoT. Les mêmes types d'échanges alléliques ont été réalisées pour le gène *arcA*, la mutation présente dans la population Ara-2 ayant été remplacée par des mutations présentes dans d'autres populations. Dans tous les cas, les capacités d'invasion de la lignée S ont été abolies. En revanche, le remplacement de l'allèle évolué du gène *gntR* de la population Ara-2 par l'allèle présent dans une autre population n'affecte pas ces capacités d'invasion. Ces expériences ont donc montré que la spécialisation écologique et les phénotypes de la lignée S étaient associés à une spécificité d'allèles au niveau des gènes *spoT* et *arcA*.

Une trajectoire génétique complexe est donc à l'origine de l'émergence de cet événement unique de diversification adaptative dans la population Ara-2. Trois gènes cibles de la sélection naturelle, des allèles spécifiques dans deux de ces gènes et un ordre précis de ces trois mutations ont permis l'émergence de la lignée S. De plus, ces trois gènes codent des régulateurs globaux de l'expression des gènes, les mutations ayant ainsi probablement des effets pleiotropes et fins sur l'expression globale des gènes dans la lignée Ara-2.

Multi-step emergence of a long-term polymorphism in *Escherichia coli*

Jessica Pluain^{1,2}, Mickaël Le Gac^{1,2†}, Olivier Tenaillon^{3,4}, Stéphane Cruveiller^{5,6}, Claudine Médigue^{5,6}, Richard E. Lenski^{7,8}, Thomas Hindré^{1,2} & Dominique Schneider^{1,2}

¹Laboratoire Adaptation et Pathogénie des Microorganismes, Université Joseph Fourier, Institut Jean Roget, F-38041 Grenoble, France. ²CNRS UMR5163, F-38041 Grenoble, France.

³INSERM, UMR-S 722, 75018 Paris, France. ⁴Université Paris Diderot, Sorbonne Paris Cité, UMR-S 722, Site Xavier Bichat, 75018 Paris, France. ⁵CNRS-UMR 8030, 91057 Evry Cedex, France. ⁶CEA/DSV/IG/Genoscope LABGeM, 91057 Evry Cedex, France. ⁷Department of Microbiology and Molecular Genetics and ⁸BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, Michigan 48824, USA. [†]Present address: Ifremer Dyneco/Pelagos, 29280 Plouzane, France.

Ecological divergence is a central process in evolution and ecology¹⁻⁷. In both experimental and natural populations under similar environmental pressures, ecological niche opportunity⁸⁻¹⁵ favoured repeated diversification of ancestral organisms into coexisting lineages. The genetic changes, underlying either niche specialization or specific phenotypic modifications that occur during these parallel divergence events, were only identified when affecting single genes or genomic regions^{8-10,16-18}. However, the genetic architecture of more complex adaptive divergence remains largely unknown^{6-8,19,20}. Here, we examined how two lineages that coexisted for >45,000 generations diverged from a common ancestor. Based on genome sequences and expression data, we identified the succession of mutational events

that enabled one lineage to invade the other. By precisely reconstructing the evolutionary history in the ancestral genetic context, we showed that three mutations were necessary and sufficient to establish the phenotypes, including negative frequency-dependent interactions, of the invading lineage. All relevant mutations affected global regulatory genes, but the singularity of the divergence resulted from the specificity of the selected alleles. At least in this system, the emergence of adaptive divergence involved a combination of historical contingency and chance.

Sympatric ecological divergence results in the emergence and maintenance of differentiated morphotypes in single environments^{3-6,9-14}. Bacterial evolution experiments not only confirmed its repeated occurrence in heterogeneous environments providing different niche opportunities and/or temporal fluctuations^{6,9,11,12}, but also showed its emergence in homogeneous simple environments. Hence, niche construction whereby living organisms alter their environment provides ecological opportunities for their own or others' success^{10,20}. Both theory and ecology/population genetics experiments have uncovered involved selective processes, including character displacement, trade-offs in life traits or metabolic abilities, cross-feeding and negative frequency-dependent selection^{3,9-12,21}. However, the occurrence of ecological divergence relies deeply on the underlying molecular determinants. Yet, except for invasion of a new niche involving a single mutational event^{10,16}, the genetic constraints weighing upon ecological divergence remain largely unexplored, thereby limiting its understanding and modelling.

Here, we identified the multiple steps and genetic constraints involved in the emergence of a unique bacterial divergence event that occurred in a long term experiment. Twelve independent populations were founded from the same strain of *Escherichia coli* and propagated in a homogeneous glucose-limited minimal medium over 55,000 generations^{22,23}. In one population

called Ara-2, two lineages, named S and L, diverged by 6,500 generations and coexisted thereafter^{20,24-26}. Divergence involves niche construction allowing invasion by the S lineage. Indeed, the L lineage grows faster on the glucose resource and produces by-products that S can better use, resulting in negative frequency-dependent selection. Global transcription profiling revealed functions specific to the S lineage that may be important for its emergence and maintenance²⁶. To characterize the genetic events triggering the emergence of this unique divergence event, we sequenced the genome of two evolved clones sampled at 6,500 generations from each S and L lineage (6.5S1, 6.5S2, 6.5L4 and 6.5L9). We identified candidate alleles and moved them in different genetic backgrounds to understand how an S-like invading genotype may emerge.

Compared to the ancestral genome²⁷, we detected an average of ~200 mutations in each of the four clones sampled from this mutator population (Supplementary Table 1), among which 68 were shared by all four clones from the S and L lineages thus being associated to the pre-divergence period (Fig. 1 and Supplementary Table 2), 55 by S clones only (Fig. 1 and Supplementary Table 3), and 36 by L clones only (Fig. 1 and Supplementary Table 4). The distribution of mutations among GO categories (Supplementary Figure 5) and the higher ratio of non-synonymous to synonymous substitution in the S lineage (Fig. 1 and Supplementary Table 6) suggested an important adaptive transition of that lineage during and after the divergence.

All the mutations were scrutinized based on the gene expression changes that were shown to be specific to each lineage²⁶. In the S lineage, 2 of the 55 mutations affected *arcA* and *gntR* encoding regulators of the TCA cycle and the Entner-Doudoroff pathway, respectively^{28,29}, that were found to be altered in that lineage²⁶. Moreover, changes in global regulatory genes are known to be important for the adaptation of the twelve long-term populations³⁰. Using our frozen

fossil record, we found both mutations in all S clones that we sampled (10 at 6,500, 1 from each 11,000, 18,000, 20,000 and 30,000 generations and 2 at 40,000 generations). Interestingly, only the evolved *arcA* allele was present in 6/18 pre-divergence 6,000-generation evolved clones showing that it was substituted before the evolved *gntR* allele.

We next constructed isogenic strains, except for one or both of the *arcA* and *gntR* alleles, in the ancestral and evolved genetic backgrounds and performed reciprocal invasion experiments with 6.5L4 as the competitor. Indeed, negative frequency-dependent interactions sustain the emergence of the S lineage and subsequent coexistence^{20,26}. While 6.5S1 invaded, when rare, 6.5L4, it lost this ability after replacing its evolved *arcA* or *gntR* allele by each ancestral counterpart (Fig. 2a and Supplementary Fig. 1a). Conversely, introducing the evolved *gntR* allele in a 6,000-generation clone (6K3) that had only the evolved *arcA* allele conferred the ability to invade 6.5L4 (Fig. 2a and Supplementary Fig. 1b). Therefore, both evolved *arcA* and *gntR* alleles are involved in the early ecological specialization of the S lineage.

When combined in the ancestral background, *arcA* and *gntR* evolved alleles did not confer the ability to invade 6.5L4 (Fig. 2b and Supplementary Fig. 2). Nevertheless, the genetically-modified ancestor was fitter when rare (Supplementary Fig. 2), suggesting ecological differences with the L lineage, but a lack of general adaptation to the evolution environment. Therefore, we added a third evolved allele which affects the *spoT* global regulatory gene in the pre-divergence period (Supplementary Table 2). Mutations in *spoT* were shown to be beneficial in the long-term populations³⁰. Combining the three evolved *spoT*, *arcA* and *gntR* alleles in the ancestral strain conferred both negative frequency-dependence and the ability to invade 6.5L4 (Fig. 2b and Supplementary Fig. 2). Therefore, the ability of the S lineage to invade the L lineage, when rare, is linked to the successive substitution of the *spoT*, *arcA* and *gntR* evolved alleles.

We next analyzed how the three alleles swept within Ara-2 and whether they conferred all the S-specific phenotypes in the ancestral background. The first mutation, *spoT*, was beneficial in the ancestor (Supplementary Fig. 3b) and the second and third, *arcA* and *gntR* respectively, were beneficial in the S lineage (Supplementary Fig. 3a). Emergence of the S lineage was therefore a multi-step process, with the three involved alleles being increasingly beneficial along the genetic path toward divergence. The evolved *arcA* and *gntR* alleles revealed different behaviors in ancestral backgrounds, suggesting pervasive epistatic interactions (Supplementary Fig. 3b). Three phenotypic abilities characterize the S lineage: maintenance over 10 days when mixed with L²⁰, growth on L-conditioned supernatant²⁰, and invasion of L clones from the same and earlier, but not later, timepoints²⁶. The three combined evolved *spoT*, *arcA* and *gntR* alleles provided the ancestor with all three phenotypes. The triple mutant maintained itself when co-cultured with 6.5L4 over 11 days (Fig. 3a), grew in the supernatant of 6.5L4 (Fig. 3b) and invaded 6.5L4 but not a mixed sample of 17,000-generation L clones (Fig. 3c).

To test the likelihood of such an ecological divergence requiring three successive mutations, we investigated why the S emergence was a unique event among the twelve evolving populations. Indeed, the long-term experiment displays a high level of genetic parallelism³⁰, and mutations were found in *spoT*, *arcA*, *arcB* and *gntR* in 7, 6, 4 and 1 other populations, respectively (Supplementary Table 7). (Both *arcA* and *arcB* genes encode the regulator and sensor of the same two-component regulatory system). Moreover, one population called Ara+6 had mutations in *spoT*, *arcB* and *gntR*. We replaced each of the three S evolved *spoT*, *arcA* and *gntR* alleles by counterparts from other populations and checked whether these changes affected the invasion ability. We constructed two strains derived from the ancestor carrying the three S evolved alleles by swapping either the *spoT* or *arcA* allele by evolved counterparts from other

populations, and four strains derived from 6.5S1: one with the *gntR* allele substituted by another evolved one, two with the *arcA* allele replaced by each of two others, and one with the ancestral *arcA* allele and an evolved *arcB* allele (Supplementary Table 7). Except for the *gntR* allele exchange, all resulting strains lost their ability to invade the L lineage (Fig. 4 and Supplementary Fig. 4). Therefore, allelic specificity in *spoT* and *arcA* was needed for ecological specialization and establishment of S-specific phenotypes.

The genetic constraints involved in the emergence of a long-term ecological divergence are still largely unknown and limit therefore our ability to understand the building up of reproductive isolation in higher organisms³¹. Using bacterial experimental evolution, we identified a complex genetic trajectory implying mutational changes, including two specific alleles, in three different genes and in a precise order of appearance that explained the emergence of a unique event of ecological divergence. This uniqueness was not related to the specificity of the genes involved in adaptation, as they were also modified in other populations, but resulted from the specificity of the selected alleles. Hence, despite convergence at the gene level, allelic diversity provided bacterial cells with tremendous phenotypic diversity, resulting in novel ecological opportunities. The relevance of this process is not restricted to asexual lineages in test tubes, but can be extended to recombining bacterial populations in nature. Indeed, ecological differentiation based on several loci was shown to gradually result in genetic isolation of *Vibrio* populations³², a process similar to those evoked in sympatric speciation in sexual eukaryotes.

Strikingly, all three mutations involved in the S ecological divergence affected global regulators of gene expression that are important for bacterial adaptation³⁰ and developmental innovations in eukaryotic cells³³. Analysis of convergence revealed that each of these genes may harbour hundreds of beneficial alleles³⁴. Hence, the combination of regulatory network plasticity and a

high diversity of beneficial alleles with contrasted impact on regulatory genes provides bacterial cells not only with the ability to cope physiologically with diverse environmental challenges but also to evolve into differentiated morphotypes able to coexist for tens of thousands of generations.

Methods Summary

We sampled two clones from each of the two S and L lineages at 6,500 generations from the polymorphic Ara-2 population of the *E. coli* long-term evolution experiment^{20,22,23}. Their genomes were resequenced as described³⁹ and mutations were identified in comparison to the ancestral genome²⁷ using a computational pipeline³⁴. Additional evolved clones were sampled throughout evolutionary time in Ara-2. Isogenic strains except for mutations of interest were constructed by allelic exchange⁴⁰ and were all derived from one of three genetic backgrounds: the ancestor, an evolved clone isolated at 6,000 generations and one of the two 6,500-generation S clones. Competition experiments were run for either 24 h (one complete serial transfer cycle) to measure fitness values and invasion abilities or 11 days with daily serial transfers to evaluate the maintenance abilities. All competitions were performed in the same medium and other conditions as used in the long-term evolution experiment. The different competitors (evolved and ancestral) were distinguished using the arabinose-utilization marker system, which allows strains to be scored by colony colour when plated on an appropriate medium and which has been shown to be selectively neutral under the conditions of the evolution experiment^{20,22}. For competitions between S-derived and L genotypes, the equivalent maltose-utilization marker system was used. The ability of strains to use by-products secreted by L clones was checked by growing them on supernatants that were collected from cultures of L clones²⁰.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Methods

Strains. All strains were sampled or derived from population Ara-2 of the *E. coli* long-term evolution experiment^{20,22-26}. In this population, two lineages, S and L, had diverged from one another by generation 6,500, and they coexisted until at least generation 40,000. Based on derived mutations specific to each lineage²⁴, we sampled two clones from each lineage from frozen samples taken at 6,500 generations during the evolution experiment: 6.5S1, 6.5S2, 6.5L4 and 6.5L9. Additional clones were sampled: 18 evolved clones during the pre-divergence period at 6,000 generations including one called 6K3, 8 S clones at 6,500 generations, 1 clone from each S and L lineage at each of 4 time points: 11,000, 18,000, 20,000 and 30,000 generations and two clones each at 40,000 generations. We also used a mixture of four L clones sampled at 17,000 generations²⁶. The Ara-2 population was founded by an ancestral strain, REL606, that cannot grow on arabinose as a carbon source^{22,27}, and all the evolved clones were also phenotypically Ara⁻. To distinguish competitors during the competition experiments, we isolated a spontaneous Ara⁺ mutant of each of 6.5S1 and 6.5L4 as described elsewhere²⁰. The arabinose-utilization marker is effectively neutral under the conditions of the long-term evolution experiment and in the S and L genetic backgrounds^{20,22}.

Genome resequencing and analyses. Genomes of clones 6.5S1, 6.5S2, 6.5L4 and 6.5L9 were resequenced on the Illumina Genome Analyzer platform using one lane of single-end 36-bp reads per genome. Candidate mutations were identified in comparison to the ancestral genome of

REL606^{27,39} using a recently developed computational pipeline³⁴. Mutations were clustered according to GO categories³⁵. The rates of synonymous and non-synonymous replacements, Ks and Ka respectively, were estimated over entire genomes using Comeron's method³⁶ implemented in libseq³⁷.

Allelic exchange. Ancestral and evolved *spoT*, *arcA*, *arcB*, *gntR* and *yqeB* alleles were moved in evolved- and/or ancestral-derived genotypes, by using the suicide plasmid pKO3⁴⁰. Briefly, PCR products containing ~500 bp of adjacent sequences on each side of each candidate allele were cloned into pKO3. The plasmids were then introduced by electro-transformation into recipients and chloramphenicol-resistant cells, formed after chromosomal integration of the plasmid, were selected. Resistant clones were streaked onto LB+sucrose plates to select for cells that lost the plasmid, which carries the *sacB* gene that makes it susceptible to killing by sucrose. These cells were then screened for the presence of the desired allele by PCR for *yqeB* and sequencing for all the other genes. Two types of controls were performed to test carefully for the presence of unintended secondary mutations that may occur during the constructions. First, each allelic exchange involving *arcA* and *gntR* was reversed to the initial allele, thereby “deconstructing” each constructed strain. Genotypic restoration was performed by the same suicide plasmid method. In all cases, the phenotypes of deconstructed strains were restored to the ones of the corresponding progenitor. Second, at the last step of sucrose selection during each allelic exchange experiment, one clone that retained the original allele was picked from the same plate from which the successful allele replacement clone was taken. The rationale was that both clones went through the same selection processes during the construction experiment, thereby likely sharing eventual secondary mutations. The control clone always had the same phenotypes as the relevant progenitor.

Competition experiments. We performed sets of pairwise competition experiments either to estimate typical fitness values or to determine whether bacteria from the same or different lineages and genetic backgrounds could invade one another when rare. Each experiment was performed under the same culture conditions used for the long-term experiment. For fitness estimation, competitors were mixed at equal cell densities, with 6 replicates for each experiment. For invasion experiments, competitors were mixed at two initial ratios (1:9 and 9:1), with 10 replicates for each pair and each initial ratio. Two types of phenotypic markers were used to distinguish the competitors. Except for experiments competing both S- and L-derived genotypes, the competitors were distinguished on the basis of an arabinose-utilization marker that is neutral (i.e., has no significant effect on fitness) under these conditions^{20,22}. For experiments involving both S- and L-derived genotypes, the competitors were distinguished on the basis of a maltose-utilization marker. Indeed, the S and L clones at 6,500 generations were found to be Mal- and Mal+, respectively. (This differential phenotype is likely related to non-synonymous mutations in *malT* that we found specifically in the S clones-Supplementary Tables 1 and 3). The 6K3 clone was also Mal- and the relevant competitions also used the maltose marker. The fitness of one competitor relative to the other is calculated as the ratio of their net growth rates during the competition^{20,22}. We performed t-tests to evaluate whether the mean fitness differed from the null hypothetical expectation of 1 for each pair of competitors at each initial ratio.

Maintenance over 11 days. The ancestor carrying the three evolved *spoT*, *arcA* and *gntR* alleles and the 6.5S1 clone were each mixed with an Ara+ revertant of 6.5L4 at two initial ratios (1:9 and 9:1), and the four mixtures were propagated for 11 days (~73 generations) under the same conditions used for the long-term experiment²⁰, with three and six replicates for 6.5S1 and the

ancestor-derived strain, respectively. The competitors were distinguished on the basis of the arabinose-utilization marker.

Growth on spent media. Filtrates of cultures of 6.5L4^{20,25} were inoculated with each of three strains: the ancestor, the ancestor carrying the three evolved *spoT*, *arcA* and *gntR* alleles, and 6.5S1, each in four replicates. Cultures were incubated at 37°C for 24 hours, and the maximum OD_{450nm} and maximum growth rate were measured.

1. Darwin, C. *On the Origin of Species by Means of Natural Selection* (Murray, 1859).
2. Mayr, E. *Populations, Species, and Evolution* (Harvard Univ. Press, 1963).
3. Schluter, D. *The Ecology of Adaptive Radiations* (Oxford Univ. Press, 2000).
4. Grant, P. R. & Grant, B. R. *How and Why Species Multiply* (Princeton Univ. Press, 2008).
5. Kondrashov, A. S. & Kondrashov, F. A. Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**, 351-354 (1999).
6. Blount, Z. D., Borland, C. Z. & Lenski R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **105**, 7899-7906 (2008).
7. Shapiro, B. J. et al. Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48-51 (2012).
8. Elmer, K. R. & Meyer A. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**, 298-306.
9. Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69-72 (1998).

10. Treves, D. S., Manning, S. & Adams, J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol. Biol. Evol.* **15**, 789-797 (1998).
11. Spencer, C. C., Bertrand, M., Travisano, M. & Doebeli, M. Adaptive diversification in genes that regulate resource use in *Escherichia coli*. *PLoS Genet.* **3**(1), e15 (2007).
12. McLean, R. C. Adaptive radiation in microbial microcosms. *J. Evol. Biol.* **18**, 1376-1386 (2005).
13. Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K. & Rodriguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115-2118 (1998).
14. Rüber, L., Verheyen, E. & Meyer, A. Replicated evolution of trophic specialization in an endemic cichlid lineage from Lake Tanganyika. *Proc. Natl Acad. Sci. USA* **96**, 10230-10235 (1999).
15. Brisson, D. & Dykhuizen, D. E. *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics* **168**, 713-722 (2004).
16. Spiers, A. J., Kahn, S. G., Bohannon, J., Travisano, M. & Rainey, P. B. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* **161**, 33-46 (2002).
17. Barrier, M., Robichaux, R. H. & Purugganan, M. D. Accelerated regulatory gene evolution in an adaptive radiation. *Proc. Natl Acad. Sci. USA* **98**, 10208-10213 (2001).
18. Albertson, R. C., Streelman, J. T. & Kocher, T. D. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proc. Natl Acad. Sci. USA* **100**, 5252-5257 (2003).
19. Fan, S., Elmer, K. R. & Meyer, A. Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and neotropical lineages. *Phil. Trans. R. Soc. B* **367**, 385-394 (2012).

20. Rozen, D. E. & Lenski, R. E. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *Am. Nat.* **155**, 24-35 (2000).
21. Doebeli, M. & Dieckmann, U. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *Am. Nat.* **156**, S77-S101 (2000).
22. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315-1341 (1991).
23. Barrick, J. E. et al. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243-1247 (2009).
24. Rozen, D. E., Schneider, D. & Lenski, R. E. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *J. Mol. Evol.* **61**, 171-180 (2005).
25. Rozen, D. E., Philippe, N., de Visser, J. A., Lenski, R. E. & Schneider, D. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecol. Lett.* **12**, 34-44 (2009).
26. Le Gac, M., Plucain, J., Hindré, T., Lenski, R. E. & Schneider, D. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **109**, 9487-9492 (2012).
27. Jeong, H. et al. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *J. Mol. Biol.* **394**, 644-652 (2009).
28. Malpica, R., Sandoval, G. R., Rodriguez, C., Franco, B. & Georgellis, D. Signaling by the arc two-component system provides a link between the redox state of the quinone pool and gene expression. *Antioxid. Redox Signal.* **8**, 781-795 (2006).
29. Peekhaus, N. & Conway, T. What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J. Bacteriol.* **180**, 3495-3502 (1998).

30. Hindré, T., Knibbe, C., Beslon, G. & Schneider, D. New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Rev. Microbiol.* **10**, 352-365 (2012).
31. Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737-741 (2009).
32. Shapiro, B. J. et al. Population genomics of early events in the ecological differentiation of bacteria. *Science* **336**, 48-51 (2012).
33. Levine, M. & Tjian, R. Transcription regulation and animal diversity. *Nature* **424**, 147-151 (2003).
34. Tenaillon, O. et al. The molecular diversity of adaptive convergence. *Science* **335**, 457-461 (2012).
35. Xu, T., Du, L. F. & Zhou, Y. Evaluation of GO-based functional similarity measures using *S. cerevisiae* protein interaction and expression profile data. *BMC Bioinformatics* **9**, 472 (2008).
36. Comeron, J. M. A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. *J. Mol. Evol.* **41**, 1152 (1995).
37. Thornton, K. Libsequence: A C++ class library for evolutionary genetic analyses. *Bioinformatics* **19**, 2325 (2003).
38. Cooper, T. F., Rozen, D. E. & Lenski, R. E. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **100**, 1072-1077 (2003).
39. Wielgoss, S. et al. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3 (Bethesda)* **1**, 183-186 (2011).
40. Link, A. J., Phillips, D. & Church, G. M. Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *J. Bacteriol.* **179**, 6228-6237 (1997).

Supplementary Information is linked to the online version of the paper at

www.nature.com/nature.

Acknowledgements This work has been supported by the Agence Nationale de la Recherche (ANR) Program “Blanc” (ANR-08-BLAN-0283-01), the Centre National de la Recherche Scientifique, and Université Joseph Fourier to D.S.; and by the US National Science Foundation (DEB-1019989) and BEACON Center for the Study of Evolution in Action to R.E.L. M.L.G. acknowledges a research fellowship from ANR. J.P. thanks the French Ministry of Research and the Université Joseph Fourier for a PhD fellowship.

Author Information Data supporting this article have been deposited to DRYAD (<http://datadryad.org/>) under the accession number XXXXX. Genome sequencing data have been deposited to ENA Short Read Archive (XXXXXX). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.S. (dominique.schneider@ujf-grenoble.fr).

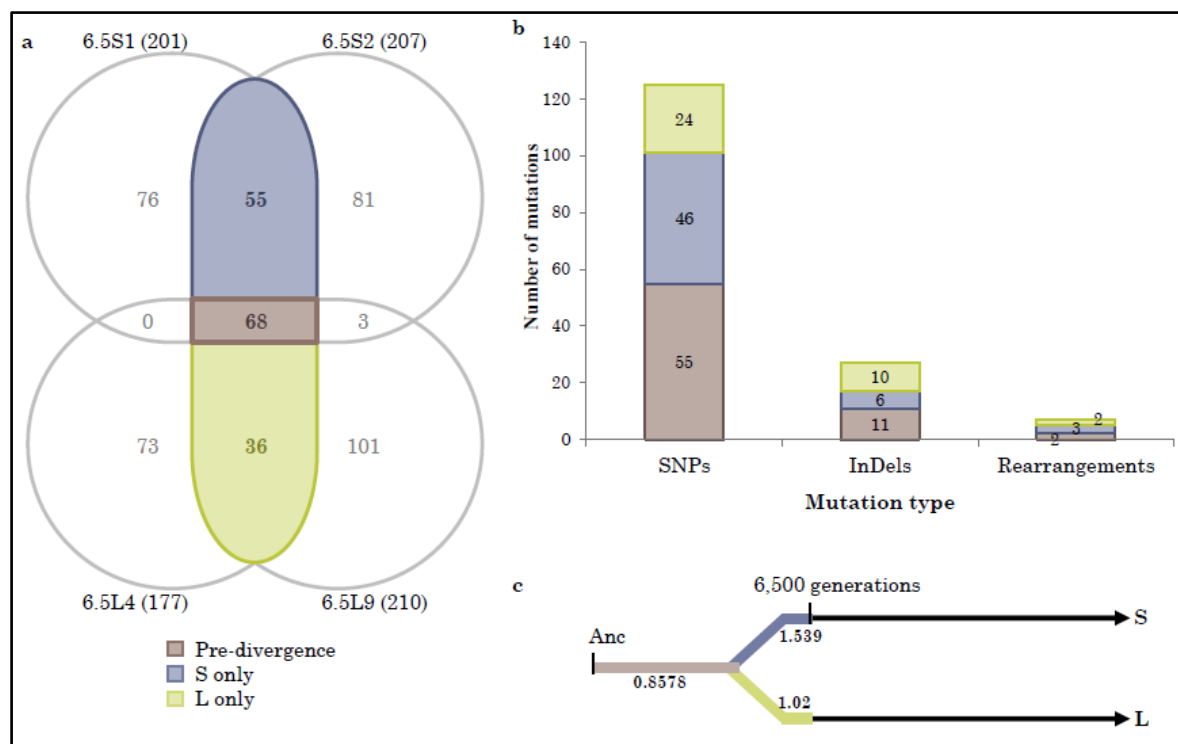


Figure 1 | Mutations identified in four early S and L genomes. a, Mutation distribution in the evolved clones 6.5S1, 6.5S2, 6.5L4 and 6.5L9. Total mutation numbers are indicated in brackets while those specific to pre-divergence, and to S and L clones only are shown in brown, blue and green zones, respectively. The numbers of all other mutations are given in grey. Not shown are two mutations shared by 6.5S1 and 6.5L9. **b,** Distribution of SNPs, InDels and rearrangements in the three mutation groups highlighted in **a**. **c,** Schematic of the three mutation groups along S and L branches. The Ka/Ks ratios are indicated for each mutation group.

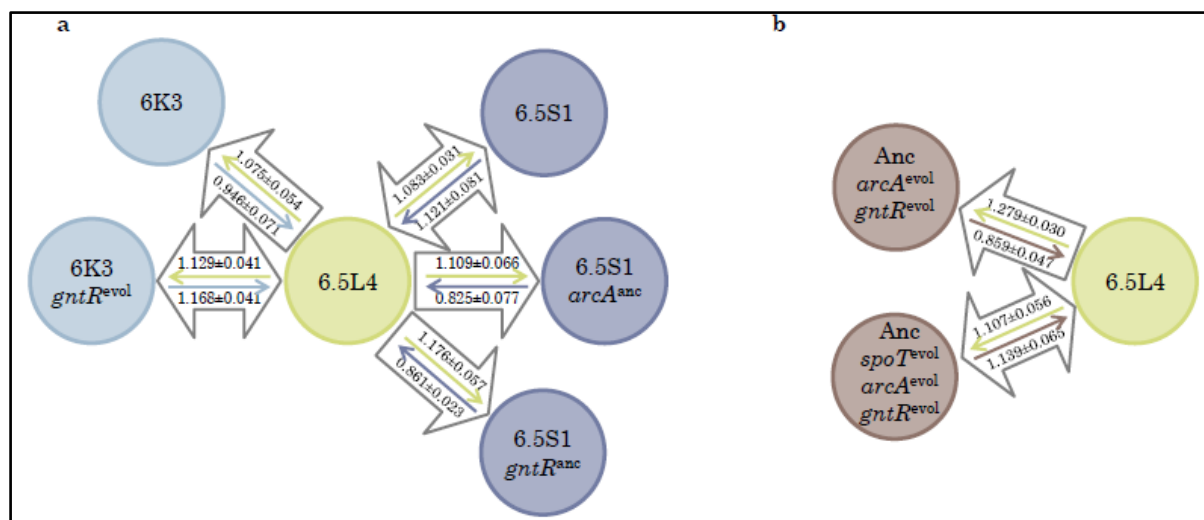


Figure 2 | Invasion experiments. **a**, Evolved genotypes sampled at 6,000 or 6,500 generations (light and dark blue, respectively) and derived isogenic strains except for *arcA* or *gntR* alleles (^{anc} for ancestral and ^{evol} for evolved) were competed against 6.5L4 (green). **b**, Ancestral strains carrying the evolved *arcA* and *gntR* alleles and either the ancestral or evolved *spoT* allele (brown) were competed against 6.5L4 (green). Small arrows indicate the competition experiments performed, with tips pointing from the initially rare (at 10% frequency) toward the initially common competitor (at 90%). Solid and dashed small arrows indicate invasion ability (fitness > 1) or inability (fitness ≤ 1), respectively, of the initially rare competitor. All small arrows are color-coded according to the initially rare competitor. Reciprocal negative frequency dependence appears as double large arrows between two competitors. Values represent the mean fitness (+/- 95% confidence interval) of the rare competitor relative to the common one. Each invasion experiment was performed with 10-fold replication.

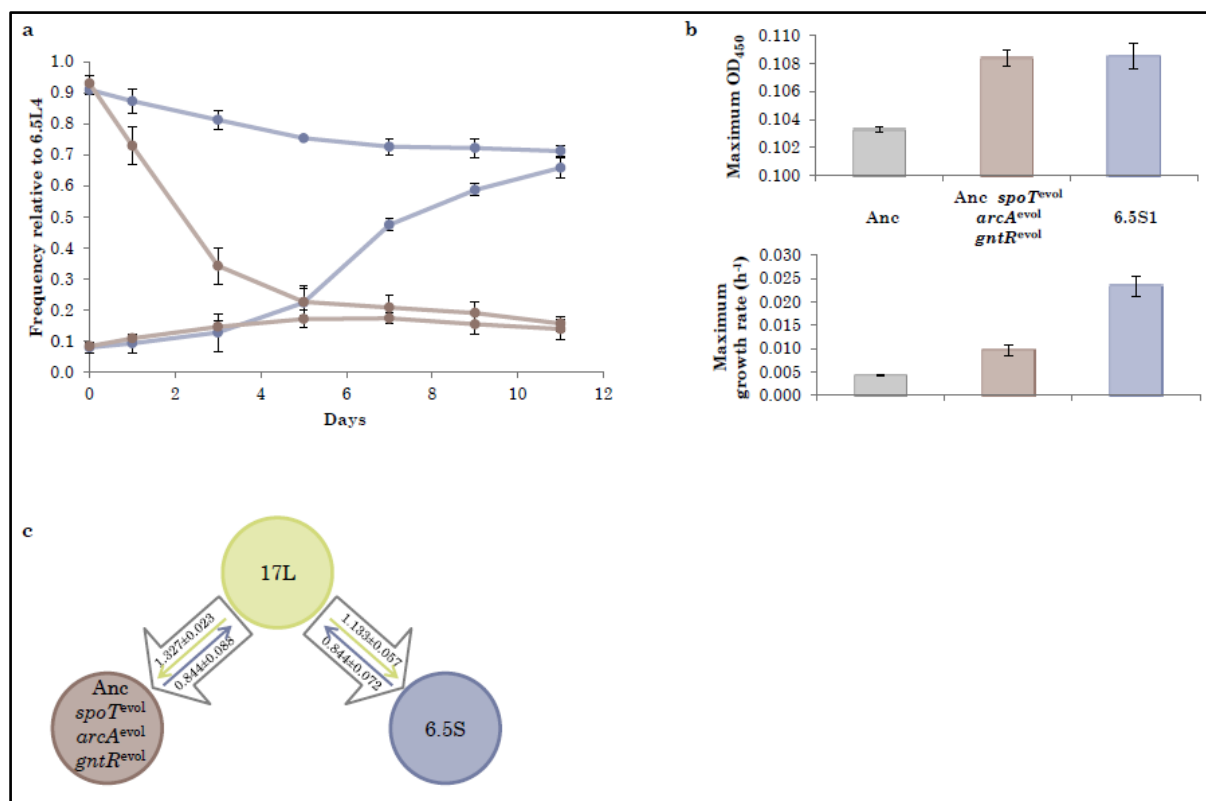


Figure 3 | Phenotypic traits of the ancestor carrying the evolved *spoT*, *arcA* and *gntR* alleles. **a**, Maintenance, at two initial ratios (10 and 90%), over an 11-day competition against 6.5L4 of the ancestor carrying the evolved *spoT*, *arcA* and *gntR* alleles and 6.5S1 (brown and blue curves, respectively). **b**, Growth parameters (maximum OD and growth rate) of the same two strains in the culture supernatant of 6.5L4. **c**, Invasion ability of the ancestor carrying the evolved *spoT*, *arcA* and *gntR* alleles and of 6,500-generation S clones (6.5S) against 17,000-generation L clones (17L). For 6.5S against 17L, data are from reference 26. Error bars represent SEs in **a,b**.

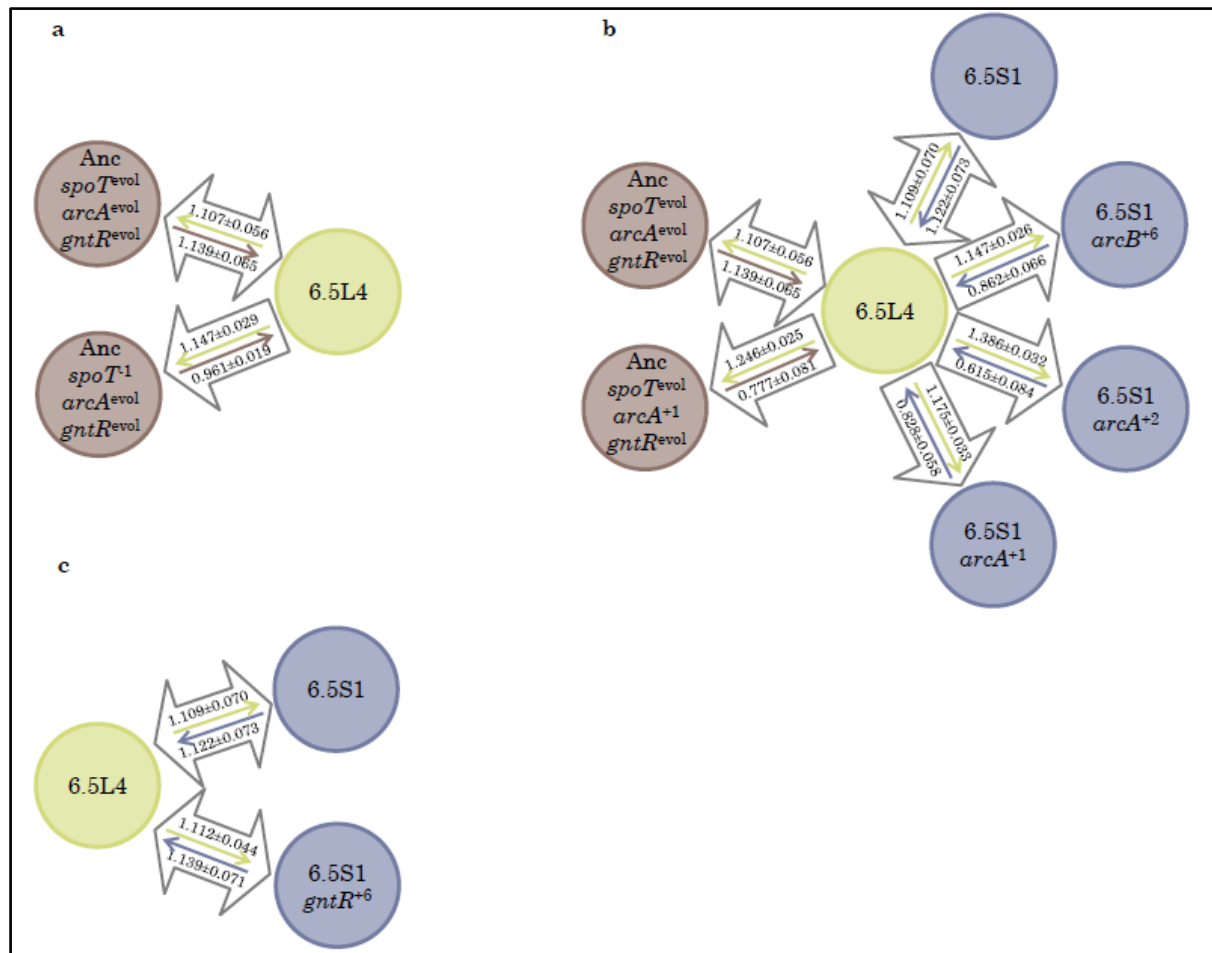
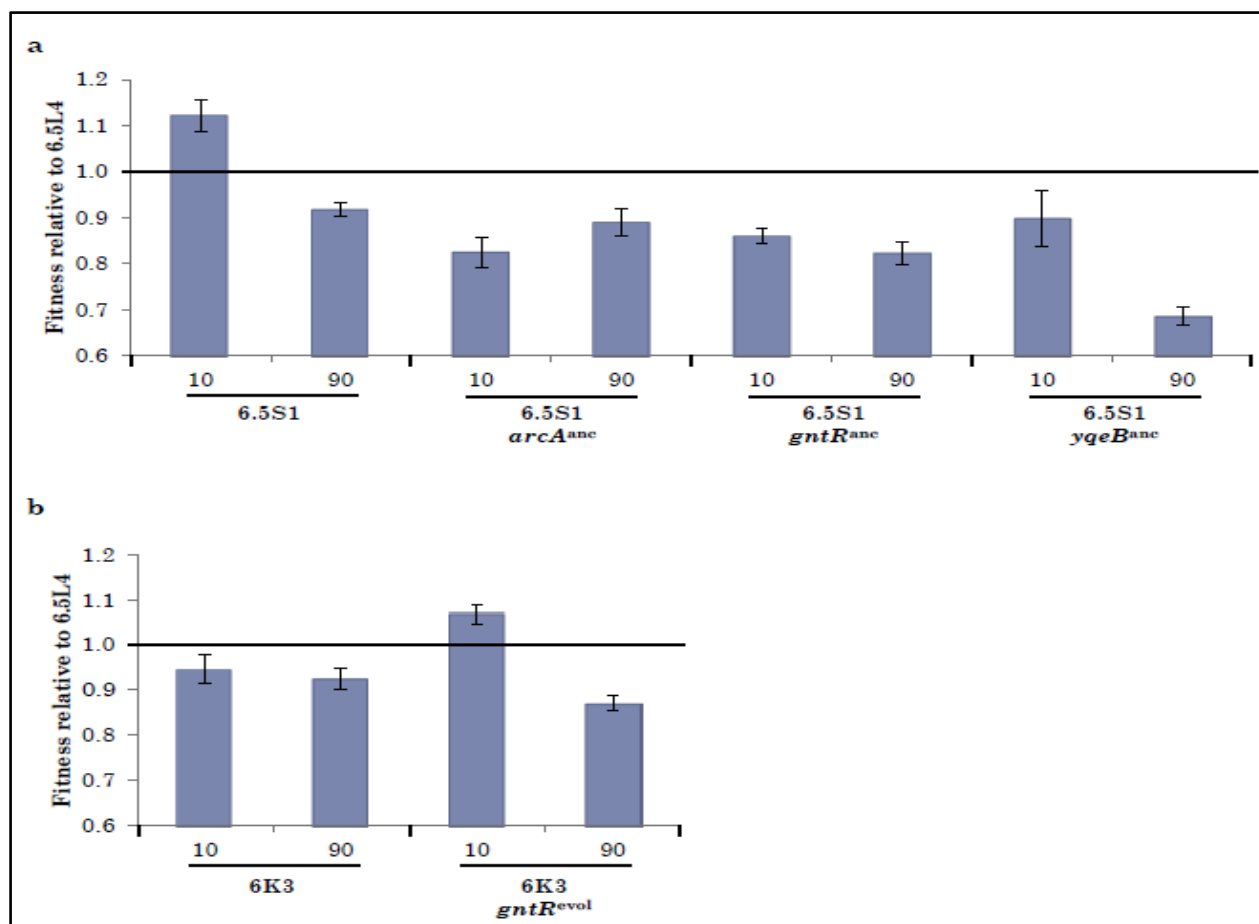


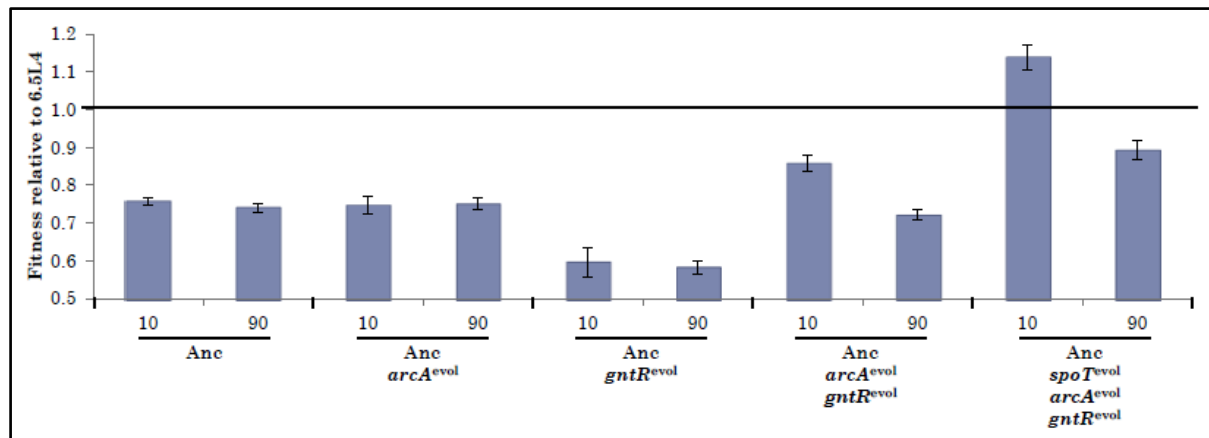
Figure 4 | Specificity of S evolved *spoT* (a), *arcA* (b) and *gntR* (c) alleles in the invasion ability of the S lineage.

The ancestor with the three S evolved alleles was changed to replace **a**, its *spoT* allele (*spoT*^{evol}) by its Ara-1 counterpart (*spoT*⁻¹); **b**, its *arcA* allele (*arcA*^{evol}) by its Ara+1 counterpart (*arcA*⁺¹), and 6.5S1 was changed by replacing its *arcA* allele by the Ara+1 (*arcA*⁺¹) and Ara+2 ones (*arcA*⁺²), and by both the ancestral *arcA* and the Ara+6 *arcB* allele (*arcB*⁺⁶). **c**, The 6.5S1 clone was changed by replacing its *gntR* allele by the Ara+6 one (*gntR*⁺⁶). Ancestor-, S- and L-derived strains are depicted in brown, blue and green circles, respectively. All invasion experiments were performed against 6.5L4, as described in Fig. 2.

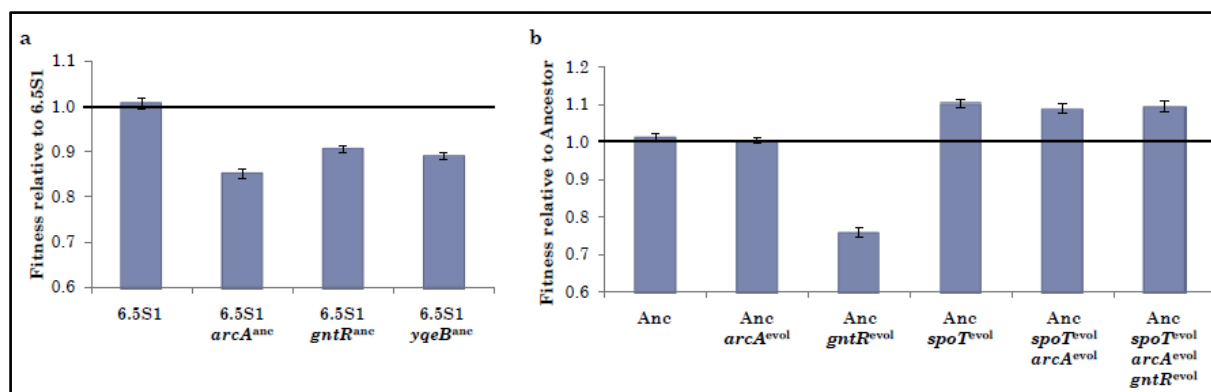


Supplementary Figure 1. Invasion properties of evolved clones and derived strains compared to the L lineage.

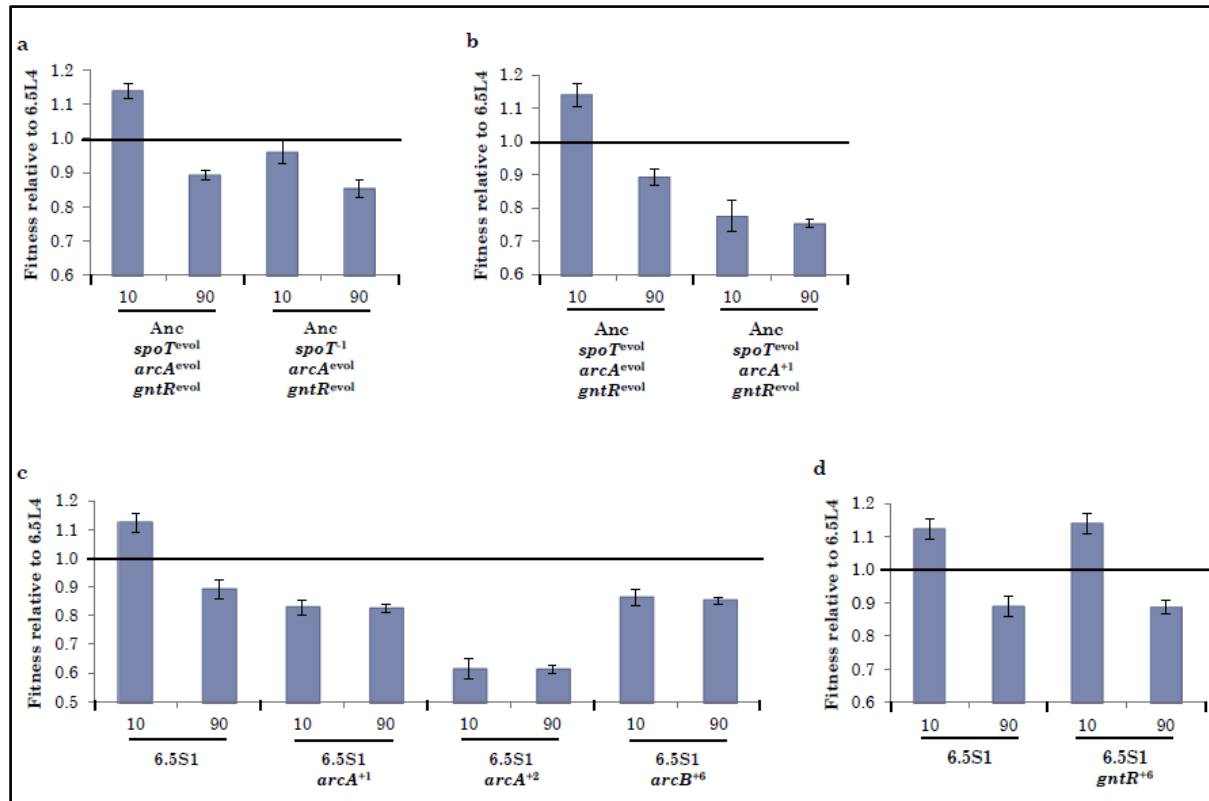
a, Fitness relative to 6.5L4 after invasion experiments with, from left to right, 6.5S1 and each of three derived strains after replacement of each of the evolved *arcA*, *gntR* and *yqeB* alleles by their ancestral counterparts. Invasion experiments were performed at each of two initial frequencies of the indicated strains (10 and 90%). The 6.5S1 clone and its derivatives are unable to use maltose as a carbon source owing to mutations in *maltT* (Supplementary Table 1; Methods) while 6.5L4 can use it. This maltose consumption trait was therefore used as a phenotypic marker to distinguish both competitors in the invasion experiments. Each relative fitness value is the mean of 10 replicate invasion experiments with error bars indicating SEs. While negative frequency-dependence was observed between 6.5S1 and 6.5L4, it was lost after replacing each of the evolved *arcA* and *gntR* alleles by the ancestral ones. This effect was specific to both alleles. Indeed, the 6.5S1-derived clone in which we replaced the evolved *yqeB* allele, which was shown to be diagnostic of S genotypes²⁴, by the ancestral allele was still much fitter when rare even if it was unable to invade 6.5L4. (This was likely due to the beneficial effect conferred by the evolved *yqeB* allele, Supplementary Fig. 3). **b**, Fitness relative to 6.5L4 after invasion experiments with either the 6,000-generation evolved 6K3 clone that carries only the evolved *arcA* allele (left) or its derived strain after introduction of the evolved *gntR* allele (right). Invasion experiments were performed as in **a**, with 6K3 being unable to use maltose as a carbon source. The introduction of the evolved *gntR* allele conferred negative frequency-dependence to 6K3.



Supplementary Figure 2. Invasion properties of ancestor-derived strains compared to the L lineage. Fitness relative to 6.5L4 after invasion experiments with, from left to right, the ancestor strain, the ancestor with the evolved *arcA* allele, the evolved *gntR* allele, both evolved alleles and the additional evolved *spoT* allele. To distinguish competitors, an Ara⁺ revertant of 6.5L4 was used in all invasion experiments. The details of invasion experiments and fitness calculations were identical to those described in Supplementary Fig. 1. While each evolved allele alone did not confer any negative frequency-dependent phenotype, combining the two evolved *arcA* and *gntR* alleles provided the ancestor with partial frequency dependence (higher fitness at 10% frequency) without invasion ability (fitness < 1 at 10% frequency). The ancestor with all three evolved *spoT*, *arcA* and *gntR* alleles displayed full negative frequency dependence, thereby mimicking an S clone.



Supplementary Figure 3. Fitness of 6.5S1- and ancestor-derived strains. **a**, Relative fitness of, from left to right, 6.5S1 and each of three derived strains after replacement of each of the evolved *arcA*, *gntR* and *yqeB* alleles by their ancestral counterparts. Each strain was used during pairwise competition experiments at 1:1 ratios against an Ara+ revertant of the 6.5S1 clone able to consume arabinose as a carbon source, which served as a phenotypic marker. Each relative fitness value is the mean of 6 replicate competition experiments with error bars indicating SEs. Each evolved allele conferred higher fitness in the 6.5S1 genetic background. **b**, Relative fitness of, from left to right, the Ara- ancestor and each of five derived strains after introduction of each of the evolved *arcA*, *gntR* and *spoT* alleles alone, of both the two evolved *spoT* and *arcA* alleles and of all three together. Each strain was used during pairwise competition experiments at 1:1 ratios against the Ara+ ancestor able to consume arabinose as a carbon source, which served as a phenotypic marker. Each relative fitness value is the mean of 6 replicate competition experiments with error bars indicating SEs. Only the evolved *spoT* allele conferred higher fitness in the ancestor, while the further successive additions of the evolved *arcA* and *gntR* alleles were neutral in this genetic context. When introduced alone in the ancestral chromosome, the evolved *arcA* and *gntR* alleles were neutral and deleterious, respectively. However, they each were beneficial in the S lineage genetic context (**a**). All these data revealed that each of the three evolved *spoT*, *arcA* and *gntR* alleles was beneficial in the respective genetic background in which they appeared during evolution of the Ara-2 population, but they also suggested pervasive epistatic interactions with other mutations present in the S lineage.



Supplementary Figure 4. Invasion properties of ancestral and evolved strains with different *spoT*, *arcA/arcB* and *gntR* alleles. **a**, Fitness relative to an Ara⁺ revertant of 6.5L4 after invasion experiments with the ancestor carrying the three evolved S *spoT*, *arcA* and *gntR* alleles (left) and the three evolved Ara-1 *spoT*, S *arcA* and S *gntR* alleles (right). The invasion ability was lost after replacement of the S evolved *spoT* allele by its Ara-1 counterpart. **b**, Fitness relative to an Ara⁺ revertant of 6.5L4 after invasion experiments with the ancestor carrying the three evolved S *spoT*, *arcA* and *gntR* alleles (left) and the three evolved S *spoT*, Ara+1 *arcA* and S *gntR* alleles (right). The invasion ability was lost after replacement of the S evolved *arcA* allele by its Ara+1 counterpart. **c**, Fitness relative to 6.5L4 after invasion experiments with, from left to right, 6.5S1 carrying the evolved *arcAB* alleles from S, Ara+1, Ara+2 and Ara+6. The invasion ability was lost after replacement of the S evolved *arcAB* alleles by the ones from the other populations. **d**, Fitness relative to 6.5L4 after invasion experiments with 6.5S1 carrying the evolved *gntR* alleles from S (left) and Ara+6 (right). The invasion ability was maintained after replacement of the S evolved *gntR* allele by its Ara+6 counterpart. All invasion experiments were performed as indicated in Supplementary Figure 1, with the arabinose and maltose phenotypic markers being used to distinguish the competitors in **a**, **b**, and **c**, **d**, respectively.

Supplementary Table 1. Mutations detected in the four evolved clones 6.5S1, 6.5S2, 6.5L4 and 6.5L9.

Clone	Position ^a	Type ^b	Base Change	Gene(s) ^c	Protein Change ^d
6.5S1	4534	NS-SNP	C->T	<i>thrC</i>	R268C
6.5S1	12074	SNP (Intergenic)	T->A	<i>yaaI</i> (87) / <i>dnaK</i> (290)	
6.5S1	14898	NS-SNP	T->C	<i>dnaJ</i>	Y245H
6.5S1	67687	S-SNP	A->G	<i>polB</i>	
6.5S1	94365	NS-SNP	C->T	<i>ftsI</i>	P50L
6.5S1	100745	S-SNP	A->G	<i>murD</i>	
6.5S1	115677	NS-SNP	A->G	<i>coaE</i>	V116A
6.5S1	178541	NS-SNP	T->C	<i>clcA</i>	L198S
6.5S1	205872	S-SNP	C->T	<i>lpxA</i>	
6.5S1	266234	S-SNP	A->G	<i>proA</i>	
6.5S1	269551	NS-SNP	C->T	<i>yagR</i>	C523Y
6.5S1	294211	Insertion (Intergenic)	1_bp	<i>ESCRE0287</i> (46) / <i>ykgE</i> (153)	
6.5S1	294649	S-SNP	A->G	<i>ykgE</i>	
6.5S1	332236	S-SNP	G->A	<i>cynS</i>	
6.5S1	332926	S-SNP	C->T	<i>cynX</i>	
6.5S1	403032	Insertion (Intergenic)	1_bp	<i>ribD</i> (37) / <i>ribE</i> (52)	
6.5S1	403037	SNP (Intergenic)	T->N	<i>ribD</i> (32) / <i>ribE</i> (57)	
6.5S1	405310	S-SNP	C->T	<i>pgpA</i>	
6.5S1	422595	Insertion (Intergenic)	1_bp	<i>cyoA</i> (180) / <i>ampG</i> (282)	
6.5S1	451466	NS-SNP	C->T	<i>maa</i>	G174S
6.5S1	459090	NS-SNP	C->T	<i>kefA</i>	T163M
6.5S1	474561	NS-SNP	A->G	<i>ybaL</i>	S249P
6.5S1	478845	Insertion	1_bp	<i>ybaK</i>	
6.5S1	478850	S-SNP	A->C	<i>ybaK</i>	
6.5S1	486377	NS-SNP	C->T	<i>cueR</i>	P73S
6.5S1	501047	Insertion	1_bp	<i>ybbD</i>	
6.5S1	506968	NS-SNP	G->A	<i>gcl</i>	R296H
6.5S1	518247	NS-SNP	C->T	<i>allD</i>	E95K
6.5S1	532468	NS-SNP	C->T	<i>sfmD</i>	R222C
6.5S1	550480	Deletion (Intergenic)	1_bp	<i>borD</i> (243) / <i>ynfO</i> (118)	
6.5S1	574371	NS-SNP	C->T	<i>cusS</i>	A435T
6.5S1	581294	NS-SNP	G->A	<i>cusA</i>	E560K
6.5S1	626325	NS-SNP	A->G	<i>rna</i>	S244P
6.5S1	649522	NS-SNP	C->T	<i>mrda</i>	G427D
6.5S1	670155	NS-SNP	A->G	<i>ECB_00624</i>	Q241R
6.5S1	677331	NS-SNP	A->G	<i>ubiF</i>	S46G
6.5S1	794715	Insertion	2_bp	<i>ECB_00735</i>	
6.5S1	804885	S-SNP	A->G	<i>bioC</i>	
6.5S1	806804	NS-SNP	A->G	<i>uvrB</i>	D218G
6.5S1	825047	SNP (Intergenic)	A->G	<i>rhIE</i> (46) / <i>ybiA</i> (186)	
6.5S1	845838	Insertion	1_bp	<i>mntR</i>	
6.5S1	850519	S-SNP	T->C	<i>ybiU</i>	
6.5S1	907400	NS-SNP	C->T	<i>grxA</i>	A21T
6.5S1	918367	S-SNP	A->G	<i>artQ</i>	

6.5S1	942704	Insertion	1_bp	<i>ESCRE0946</i>	
6.5S1	950771	NS-SNP	A->G	<i>ftsK</i>	T164A
6.5S1	1030571	S-SNP	C->T	<i>pqiB</i>	
6.5S1	1037171	SNP (Intergenic)	T->C	<i>ompA</i> (330) / <i>sulA</i> (27)	
6.5S1	1037686	NS-SNP	T->C	<i>sulA</i>	T109A
6.5S1	1042350	S-SNP	A->G	<i>helD</i>	
6.5S1	1046824	Deletion	1_bp	<i>yccW</i>	
6.5S1	1053644	NS-SNP	A->G	<i>hyaE</i>	D67G
6.5S1	1057125	NS-SNP	G->A	<i>appB</i>	V247M
6.5S1	1094105	S-SNP	T->C	<i>putA</i>	
6.5S1	1149462	S-SNP	G->A	<i>flgG</i>	
6.5S1	1159018	SNP (Intergenic)	T->C	<i>rne</i> (53) / <i>yceQ</i> (82)	
6.5S1	1212789	SNP (Intergenic)	A->G	<i>elbA</i> (75) / <i>ycgX</i> (621)	
6.5S1	1213508	NS-SNP	A->G	<i>ycgE</i>	V238A
6.5S1	1246849	S-SNP	C->T	<i>treA</i>	
6.5S1	1333848	SNP (Intergenic)	A->G	<i>ECB_01252</i> (119) / <i>ESCRE1333</i> (18)	
6.5S1	1354410	NS-SNP	G->A	<i>sapA</i>	T357I
6.5S1	1355035	NS-SNP	C->T	<i>sapA</i>	D149N
6.5S1	1363126	NS-SNP	A->G	<i>puuB</i>	H176R
6.5S1	1383106	S-SNP	G->A	<i>ycjX</i>	
6.5S1	1413037	NS-SNP	G->A	<i>recE</i>	P507L
6.5S1	1447371	S-SNP	T->C	<i>ynbA</i>	
6.5S1	1453944	NS-SNP	C->T	<i>hrpA</i>	A499V
6.5S1	1465797	NS-SNP	A->G	<i>ycdJ</i>	E134G
6.5S1	1524453	NS-SNP	A->G	<i>ECB_01432</i>	E968G
6.5S1	1532723	NS-SNP	A->G	<i>ddpD</i>	V78A
6.5S1	1580589	S-SNP	A->G	<i>yneI</i>	
6.5S1	1614900	IS_Insertion	IS3	<i>ECB_01523</i>	
6.5S1	1646650	S-SNP	A->G	<i>ynfL</i>	
6.5S1	1689708	Insertion (Intergenic)	1_bp	<i>nth</i> (255) / <i>ydgR</i> (356)	
6.5S1	1695430	NS-SNP	T->C	<i>ydhA</i>	S54G
6.5S1	1711520	Insertion (Intergenic)	1_bp	<i>ydhD</i> (109) / <i>ydhO</i> (225)	
6.5S1	1733754	NS-SNP	A->T	<i>pykF</i>	I264F
6.5S1	1741769	SNP (Intergenic)	T->A	<i>sufA</i> (211) / <i>ECB_misc_RNA_31</i> (116)	
6.5S1	1815710	NS-SNP	T->C	<i>ynjC</i>	V409A
6.5S1	1826622	S-SNP	T->C	<i>sppA</i>	
6.5S1	1868853	S-SNP	T->C	<i>yeaY</i>	
6.5S1	1940852	Deletion	1_bp	<i>yecT</i>	
6.5S1	1944477	Deletion	1_bp	<i>flhB</i>	
6.5S1	1944481	NS-SNP	A->C	<i>flhB</i>	F161V
6.5S1	1969213	Deletion	1_bp	<i>yecA</i>	
6.5S1	1974150	NS-SNP	A->G	<i>hchA</i>	Y42C
6.5S1	2007562	Insertion	1_bp	<i>ECB_01906</i>	
6.5S1	2010661	NS-SNP	A->G	<i>yeeA</i>	V125A
6.5S1	2014169	NS-SNP	G->A	<i>sbcB</i>	G324E
6.5S1	2015168	S-SNP	T->C	<i>yeeE</i>	
6.5S1	2088363	NS-SNP	C->T	<i>mdtA</i>	R414C
6.5S1	2094204	NS-SNP	T->C	<i>mdtC</i>	L904P
6.5S1	2107226	Insertion	1_bp	<i>ECB_01994</i>	
6.5S1	2111272	Insertion	1_bp	<i>ECB_01999</i>	

6.5S1	2116042	Insertion	1_bp	<i>ECB_02003</i>	
6.5S1	2128584	NS-SNP	C->T	<i>gatZ</i>	M10I
6.5S1	2169035	S-SNP	C->T	<i>yehY</i>	
6.5S1	2171395	S-SNP	G->A	<i>bglX</i>	
6.5S1	2175382	NS-SNP	C->T	<i>pbpG</i>	V243M
6.5S1	2186143	S-SNP	T->C	<i>yeiT</i>	
6.5S1	2213468	NS-SNP	T->C	<i>fruK</i>	N29S
6.5S1	2216363	NS-SNP	G->A	<i>yeiW</i>	R41C
6.5S1	2263557	SNP (Intergenic)	T->C	<i>IS1_insA(170) / rcsC(105)</i>	
6.5S1	2322337	IS_Insertion	IS186	<i>menC</i>	
6.5S1	2378672	NS-SNP	T->C	<i>truA</i>	H186R
6.5S1	2409014	NS-SNP	G->A	<i>vacJ</i>	P146S
6.5S1	2413069	S-SNP	A->G	<i>dsdX</i>	
6.5S1	2418405	NS-SNP	G->A	<i>evgA</i>	D142N
6.5S1	2431702	NS-SNP	T->C	<i>yfdZ</i>	H273R
6.5S1	2442609	SNP (Intergenic)	A->G	<i>ypdH(77) / glk(142)</i>	
6.5S1	2452949	S-SNP	G->A	<i>gltX</i>	
6.5S1	2476458	S-SNP	C->T	<i>murP</i>	
6.5S1	2522405	NS-SNP	T->C	<i>hyfA</i>	V126A
6.5S1	2616297	Deletion (Intergenic)	1_bp	<i>purL(359) / yfhD(37)</i>	
6.5S1	2633002	NS-SNP	C->T	<i>nadB</i>	R364C
6.5S1	2633980	NS-SNP	G->A	<i>yfiC</i>	P93L
6.5S1	2639803	S-SNP	G->A	<i>yfiF</i>	
6.5S1	2646652	S-SNP	T->C	<i>kgtP</i>	
6.5S1	2707687	NS-SNP	G->A	<i>emrB</i>	W71*
6.5S1	2719798	NS-SNP	A->G	<i>mltB</i>	L213P
6.5S1	2722755	NS-SNP	A->G	<i>srlD</i>	T54A
6.5S1	2735375	NS-SNP	T->C	<i>ascF</i>	M331T
6.5S1	2784199	S-SNP	A->G	<i>ygcQ</i>	
6.5S1	2790136	NS-SNP	T->C	<i>yqcE</i>	S349P
6.5S1	2798243	NS-SNP	G->A	<i>pyrG</i>	H118Y
6.5S1	2800862	NS-SNP	T->C	<i>relA</i>	D573G
6.5S1	2814511	NS-SNP	G->A	<i>queF</i>	E79K
6.5S1	2836224	S-SNP	G->A	<i>ygdD</i>	
6.5S1	2837475	SNP (Intergenic)	T->C	<i>ECB_misc_RNA_9(62) / ygdI(14)</i>	
6.5S1	2850146	S-SNP	T->C	<i>recB</i>	
6.5S1	2897437	NS-SNP	G->A	<i>yqeA</i>	M74I
6.5S1	2899613	IS_Insertion	IS150	<i>yqeB</i>	
6.5S1	2976544	S-SNP	T->C	<i>endA</i>	
6.5S1	2977749	NS-SNP	G->A	<i>gshB</i>	G41S
6.5S1	2997456	Deletion	1_bp	<i>ECB_02798</i>	
6.5S1	3008434	Insertion (Intergenic)	1_bp	<i>yeeT(403) / yeeU(309)</i>	
6.5S1	3015347	S-SNP	C->T	<i>kpsD</i>	
6.5S1	3037784	NS-SNP	G->A	<i>epsD</i>	T404M
6.5S1	3057696	S-SNP	T->C	<i>ECB_02851</i>	
6.5S1	3064150	NS-SNP	T->C	<i>ECB_02857</i>	N54S
6.5S1	3067566	NS-SNP	T->C	<i>yghS</i>	K160R
6.5S1	3075592	S-SNP	G->A	<i>hybC</i>	
6.5S1	3123048	NS-SNP	A->G	<i>yqiH</i>	Y111C
6.5S1	3123349	S-SNP	T->C	<i>yqiH</i>	

6.5S1	3158462	NS-SNP	G->A	<i>ebgA</i>	A1005T
6.5S1	3162869	NS-SNP	G->A	<i>ygjK</i>	V389I
6.5S1	3168389	NS-SNP	C->T	<i>ygjO</i>	D102N
6.5S1	3179772	S-SNP	T->C	<i>exuR</i>	
6.5S1	3203410	NS-SNP	T->C	<i>ECB_misc_RNA_46</i>	K18E
6.5S1	3218529	NS-SNP	G->A	<i>kbaY</i>	G52R
6.5S1	3220182	NS-SNP	A->G	<i>agaC</i>	T89A
6.5S1	3220975	NS-SNP	G->A	<i>agaD</i>	D89N
6.5S1	3250447	NS-SNP	G->A	<i>infB</i>	H268Y
6.5S1	3256388	NS-SNP	T->C	<i>yhbX</i>	S154G
6.5S1	3343907	S-SNP	C->T	<i>acrF</i>	
6.5S1	3385583	S-SNP	G->A	<i>gspD</i>	
6.5S1	3419452	NS-SNP	C->T	<i>yhfG</i>	R25H
6.5S1	3425074	NS-SNP	T->C	<i>nirC</i>	V56A
6.5S1	3478428	S-SNP	G->A	<i>malQ</i>	
6.5S1	3483399	NS-SNP	T->C	<i>malT</i>	L572P
6.5S1	3509731	NS-SNP	T->C	<i>gntR</i>	D150G
6.5S1	3511189	Deletion (Intergenic)	1_bp	<i>yhhW(32) / yhhX(91)</i>	
6.5S1	3512018	NS-SNP	A->T	<i>yhhX</i>	Y81N
6.5S1	3514265	NS-SNP	T->C	<i>yhhZ</i>	V317A
6.5S1	3566519	S-SNP	A->G	<i>yhiM</i>	
6.5S1	3610949	NS-SNP	T->C	<i>yhjJ</i>	S162G
6.5S1	3670123	SNP (Intergenic)	T->C	<i>malS(38) / avtA(140)</i>	
6.5S1	3677560	S-SNP	G->A	<i>lyxK</i>	
6.5S1	3762120	NS-SNP	C->A	<i>spoT</i>	A455D
6.5S1	3779088	NS-SNP	G->T	<i>ECB_03519</i>	A234D
6.5S1	3866000	Deletion (Intergenic)	1_bp	<i>bglF(98) / bglG(35)</i>	
6.5S1	3891539	Insertion	1_bp	<i>ESCRE3877</i>	
6.5S1	3894997	LargeDeletion	4900_bp	<i>rbsD / rbsK</i>	
6.5S1	3899899	IS_Insertion	IS150	<i>rbsK</i>	
6.5S1	3925659	NS-SNP	C->T	<i>gpp</i>	A125T
6.5S1	3985793	NS-SNP	T->C	<i>ECB_03714</i>	T201A
6.5S1	4085858	Insertion (Intergenic)	1_bp	<i>cpxP(93) / fieF(56)</i>	
6.5S1	4101025	NS-SNP	C->T	<i>hslU</i>	R69H
6.5S1	4117175	NS-SNP	A->G	<i>yjF</i>	V197A
6.5S1	4185358	S-SNP	G->A	<i>purD</i>	
6.5S1	4202021	NS-SNP	C->T	<i>iclR</i>	R180H
6.5S1	4218174	Insertion	1_bp	<i>yjbH</i>	
6.5S1	4218862	Insertion	1_bp	<i>ESCRE4211</i>	
6.5S1	4225823	S-SNP	A->G	<i>malK</i>	
6.5S1	4235048	S-SNP	A->G	<i>plsB</i>	
6.5S1	4266390	SNP (Intergenic)	C->A	<i>acs(307) / nrfA(86)</i>	
6.5S1	4303399	NS-SNP	G->A	<i>phnC</i>	T194I
6.5S1	4331214	Deletion (Intergenic)	1_bp	<i>yjdJ(90) / yjdK(137)</i>	
6.5S1	4334407	NS-SNP	A->G	<i>yjdL</i>	V242A
6.5S1	4357593	Insertion	1_bp	<i>ampC</i>	
6.5S1	4361415	NS-SNP	C->T	<i>poxA</i>	A21V
6.5S1	4375768	Insertion	6_bp	<i>mutL</i>	
6.5S1	4425769	S-SNP	A->G	<i>msrA</i>	
6.5S1	4471801	Insertion	1_bp	<i>yjgR</i>	

6.5S1	4504339	SNP (Intergenic)	T->A	IS1_insB(293) / ESCRE4497(83)	
6.5S1	4506111	LargeDeletion	15,458_bp	sgcR / fimB	
6.5S1	4584759	NS-SNP	A->G	yjiM	F13L
6.5S1	4590035	NS-SNP	G->A	dnaC	R59C
6.5S1	4591454	S-SNP	G->A	yjiP	
6.5S1	4615700	NS-SNP	G->A	nadR	E58K
6.5S1	4628190	NS-SNP	G->A	arcA	L93F
6.5S2	12074	SNP (Intergenic)	T->A	yaaI(87) / dnaK(290)	
6.5S2	94365	NS-SNP	C->T	ftsI	P50L
6.5S2	100745	S-SNP	A->G	murD	
6.5S2	112497	NS-SNP	C->T	secA	A472V
6.5S2	115677	NS-SNP	A->G	coaE	V116A
6.5S2	141859	NS-SNP	A->G	gcd	F724L
6.5S2	171587	NS-SNP	A->G	thiA	N421S
6.5S2	178541	NS-SNP	T->C	clcA	L198S
6.5S2	189403	NS-SNP	A->G	glnD	V697A
6.5S2	205872	S-SNP	C->T	lpxA	
6.5S2	235039	NS-SNP	A->G	yafE	I64V
6.5S2	266234	S-SNP	A->G	proA	
6.5S2	269551	NS-SNP	C->T	yagR	C523Y
6.5S2	294208	Insertion (Intergenic)	1_bp	ESCRE0287(49) / ykgE(150)	
6.5S2	294649	S-SNP	A->G	ykgE	
6.5S2	332236	S-SNP	G->A	cynS	
6.5S2	332926	S-SNP	C->T	cynX	
6.5S2	343533	NS-SNP	A->G	mhpB	N234S
6.5S2	352371	Insertion (Intergenic)	1_bp	frmR(53) / yaiO(133)	
6.5S2	451466	NS-SNP	C->T	maa	G174S
6.5S2	474561	NS-SNP	A->G	ybaL	S249P
6.5S2	479515	S-SNP	C->T	ybaP	
6.5S2	486377	NS-SNP	C->T	cueR	P73S
6.5S2	501047	Insertion	1_bp	ybbD	
6.5S2	501053	NS-SNP	C->T	ybbD	L83F
6.5S2	505920	S-SNP	C->T	allR	
6.5S2	506968	NS-SNP	G->A	gcl	R296H
6.5S2	574371	NS-SNP	C->T	cusS	A435T
6.5S2	604650	NS-SNP	G->A	fepD	T83I
6.5S2	626325	NS-SNP	A->G	rna	S244P
6.5S2	649522	NS-SNP	C->T	mrda	G427D
6.5S2	777377	S-SNP	C->T	modC	
6.5S2	794715	Insertion	1_bp	ECB_00735	
6.5S2	795317	Insertion (Intergenic)	1_bp	ECB_00736(20) / ECB_00737(22)	
6.5S2	804885	S-SNP	A->G	bioC	
6.5S2	806804	NS-SNP	A->G	uvrB	D218G
6.5S2	825047	SNP (Intergenic)	A->G	rhIE(46) / ybiA(186)	
6.5S2	845838	Insertion	1_bp	mntR	
6.5S2	864245	NS-SNP	C->T	yliC	L243F
6.5S2	876483	NS-SNP	G->A	cmr	A87T
6.5S2	932214	NS-SNP	T->C	aqpZ	T181A
6.5S2	950771	NS-SNP	A->G	ftsK	T164A
6.5S2	990830	NS-SNP	C->T	smtA	Q68*

6.5S2	1002129	NS-SNP	C->T	<i>aspC</i>	A225T
6.5S2	1018108	Deletion	1_bp	<i>ycbS</i>	
6.5S2	1018114	NS-SNP	A->T	<i>ycbS</i>	N603I
6.5S2	1025721	NS-SNP	A->G	<i>ycbY</i>	S263G
6.5S2	1030571	S-SNP	C->T	<i>pqiB</i>	
6.5S2	1053644	NS-SNP	A->G	<i>hyaE</i>	D67G
6.5S2	1057125	NS-SNP	G->A	<i>appB</i>	V247M
6.5S2	1149462	S-SNP	G->A	<i>flgG</i>	
6.5S2	1159018	SNP (Intergenic)	T->C	<i>rne</i> (53) / <i>yceQ</i> (82)	
6.5S2	1162166	SNP (Intergenic)	A->G	<i>rpmF</i> (24) / <i>plsX</i> (57)	
6.5S2	1246849	S-SNP	C->T	<i>treA</i>	
6.5S2	1313490	NS-SNP	C->T	<i>yciG</i>	G47D
6.5S2	1332485	NS-SNP	C->T	<i>cysB</i>	H87Y
6.5S2	1349560	Insertion (Intergenic)	1_bp	<i>fabI</i> (17) / <i>ESCRE1348</i> (152)	
6.5S2	1354410	NS-SNP	G->A	<i>sapA</i>	T357I
6.5S2	1383106	S-SNP	G->A	<i>ycjX</i>	
6.5S2	1462266	IS_Insertion (Intergenic)	IS150	<i>mokB</i> (313) / <i>trg</i> (28)	
6.5S2	1465797	NS-SNP	A->G	<i>ydcJ</i>	E134G
6.5S2	1480629	NS-SNP	A->G	<i>ydcR</i>	T173A
6.5S2	1600163	NS-SNP	G->A	<i>ESCRE1593</i>	G29R
6.5S2	1613808	S-SNP	A->G	<i>ECB_01522</i>	
6.5S2	1614900	IS_Insertion	IS3	<i>ECB_01523</i>	
6.5S2	1646650	S-SNP	A->G	<i>ynfL</i>	
6.5S2	1689708	Insertion (Intergenic)	1_bp	<i>nth</i> (255) / <i>ydgR</i> (356)	
6.5S2	1709068	NS-SNP	G->A	<i>lhr</i>	A930T
6.5S2	1711520	Insertion (Intergenic)	1_bp	<i>ydhD</i> (109) / <i>ydhO</i> (225)	
6.5S2	1733754	NS-SNP	A->T	<i>pykF</i>	I264F
6.5S2	1742137	SNP (Intergenic)	G->A	<i>ECB_misc_RNA_31</i> (64) / <i>ydiH</i> (90)	
6.5S2	1815710	NS-SNP	T->C	<i>ynjC</i>	V409A
6.5S2	1826622	S-SNP	T->C	<i>sppA</i>	
6.5S2	1858729	S-SNP	G->A	<i>yeaS</i>	
6.5S2	1923734	S-SNP	T->C	<i>ruvB</i>	
6.5S2	1940852	Deletion	1_bp	<i>yecT</i>	
6.5S2	1944477	Deletion	1_bp	<i>flhB</i>	
6.5S2	1969213	Deletion	1_bp	<i>yecA</i>	
6.5S2	1996748	NS-SNP	C->A	<i>yeeO</i>	A367S
6.5S2	2094204	NS-SNP	T->C	<i>mdtC</i>	L904P
6.5S2	2097904	NS-SNP	A->G	<i>baeR</i>	H174R
6.5S2	2111272	Insertion	1_bp	<i>ECB_01999</i>	
6.5S2	2131459	Deletion (Intergenic)	1_bp	<i>gatY</i> (182) / <i>fbaB</i> (118)	
6.5S2	2169035	S-SNP	C->T	<i>yehY</i>	
6.5S2	2171395	S-SNP	G->A	<i>bglX</i>	
6.5S2	2175382	NS-SNP	C->T	<i>pbpG</i>	V243M
6.5S2	2186143	S-SNP	T->C	<i>yeiT</i>	
6.5S2	2213468	NS-SNP	T->C	<i>fruK</i>	N29S
6.5S2	2216363	NS-SNP	G->A	<i>yeiW</i>	R41C
6.5S2	2260988	S-SNP	G->A	<i>yojL</i>	
6.5S2	2277650	S-SNP	T->C	<i>ECB_02154</i>	
6.5S2	2322347	Insertion	1_bp	<i>menC</i>	
6.5S2	2356846	SNP (Intergenic)	G->A	<i>yfbV</i> (216) / <i>ackA</i> (123)	

6.5S2	2369094	NS-SNP	G->A	<i>hisQ</i>	R139W
6.5S2	2378672	NS-SNP	T->C	<i>truA</i>	H186R
6.5S2	2384370	S-SNP	G->A	<i>fabB</i>	
6.5S2	2411665	S-SNP	A->G	<i>dsdC</i>	
6.5S2	2418405	NS-SNP	G->A	<i>evgA</i>	D142N
6.5S2	2431702	NS-SNP	T->C	<i>yfdZ</i>	H273R
6.5S2	2442609	SNP (Intergenic)	A->G	<i>ypdH(77) / glk(142)</i>	
6.5S2	2452949	S-SNP	G->A	<i>gltX</i>	
6.5S2	2520683	Insertion (Intergenic)	1_bp	<i>dapA(51) / gcvR(95)</i>	
6.5S2	2522405	NS-SNP	T->C	<i>hyfA</i>	V126A
6.5S2	2626060	NS-SNP	A->G	<i>lepB</i>	V248A
6.5S2	2633002	NS-SNP	C->T	<i>nadB</i>	R364C
6.5S2	2633980	NS-SNP	G->A	<i>yfiC</i>	P93L
6.5S2	2639803	S-SNP	G->A	<i>yfiF</i>	
6.5S2	2646652	S-SNP	T->C	<i>kgtP</i>	
6.5S2	2710931	NS-SNP	T->C	<i>gshA</i>	N123S
6.5S2	2735375	NS-SNP	T->C	<i>ascF</i>	M331T
6.5S2	2772516	Deletion (Intergenic)	1_bp	<i>iap(395) / IS186_insL(37)</i>	
6.5S2	2784199	S-SNP	A->G	<i>ygcQ</i>	
6.5S2	2798243	NS-SNP	G->A	<i>pyrG</i>	H118Y
6.5S2	2800862	NS-SNP	T->C	<i>relA</i>	D573G
6.5S2	2811261	SNP (Intergenic)	G->A	<i>gudP(203) / yqcA(232)</i>	
6.5S2	2830150	SNP (Intergenic)	G->A	<i>ECB_02652(53) / fucI(417)</i>	
6.5S2	2842327	Insertion (Intergenic)	1_bp	<i>metV(49) / amiC(25)</i>	
6.5S2	2897437	NS-SNP	G->A	<i>yqeA</i>	M74I
6.5S2	2899613	IS_Insertion	IS150	<i>yqeB</i>	
6.5S2	2976544	S-SNP	T->C	<i>endA</i>	
6.5S2	2977749	NS-SNP	G->A	<i>gshB</i>	G41S
6.5S2	3037784	NS-SNP	G->A	<i>epsD</i>	T404M
6.5S2	3046970	Insertion	1_bp	<i>yghK</i>	
6.5S2	3048591	NS-SNP	G->A	<i>glcB</i>	Q703*
6.5S2	3064150	NS-SNP	T->C	<i>ECB_02857</i>	N54S
6.5S2	3067566	NS-SNP	T->C	<i>yghS</i>	K160R
6.5S2	3075592	S-SNP	G->A	<i>hybC</i>	
6.5S2	3123048	NS-SNP	A->G	<i>yqiH</i>	Y111C
6.5S2	3129096	S-SNP	G->A	<i>rfaE</i>	
6.5S2	3158462	NS-SNP	G->A	<i>ebgA</i>	A1005T
6.5S2	3162869	NS-SNP	G->A	<i>ygjK</i>	V389I
6.5S2	3168389	NS-SNP	C->T	<i>ygjO</i>	D102N
6.5S2	3191269	Deletion (Intergenic)	1_bp	<i>yhaO(50) / tdcG(289)</i>	
6.5S2	3203410	NS-SNP	T->C	<i>ECB_misc_RNA_46</i>	K18E
6.5S2	3218529	NS-SNP	G->A	<i>kbaY</i>	G52R
6.5S2	3220182	NS-SNP	A->G	<i>agaC</i>	T89A
6.5S2	3220975	NS-SNP	G->A	<i>agaD</i>	D89N
6.5S2	3246305	NS-SNP	A->G	<i>pnp</i>	V56A
6.5S2	3250447	NS-SNP	G->A	<i>infB</i>	H268Y
6.5S2	3252240	NS-SNP	C->T	<i>nusA</i>	R174H
6.5S2	3256388	NS-SNP	T->C	<i>yhbX</i>	S154G
6.5S2	3304164	SNP (Intergenic)	G->A	<i>dcuD(13) / sspB(30)</i>	
6.5S2	3338589	NS-SNP	T->C	<i>dusB</i>	I140T

6.5S2	3343907	S-SNP	C->T	<i>acrF</i>	
6.5S2	3351179	NS-SNP	G->N	<i>ESCRE3333</i>	A79R
6.5S2	3354270	Insertion	1_bp	<i>rrlD</i>	
6.5S2	3385583	S-SNP	G->A	<i>gspD</i>	
6.5S2	3419452	NS-SNP	C->T	<i>yhfG</i>	R25H
6.5S2	3425074	NS-SNP	T->C	<i>nirC</i>	V56A
6.5S2	3478428	S-SNP	G->A	<i>malQ</i>	
6.5S2	3483050	NS-SNP	G->A	<i>malT</i>	A456T
6.5S2	3509731	NS-SNP	T->C	<i>gntR</i>	D150G
6.5S2	3512018	NS-SNP	A->T	<i>yhhX</i>	Y81N
6.5S2	3514265	NS-SNP	T->C	<i>yhhZ</i>	V317A
6.5S2	3566519	S-SNP	A->G	<i>yhiM</i>	
6.5S2	3600935	NS-SNP	G->A	<i>yhjB</i>	T151I
6.5S2	3610949	NS-SNP	T->C	<i>yhjJ</i>	S162G
6.5S2	3670123	SNP (Intergenic)	T->C	<i>malS</i> (38) / <i>avtA</i> (140)	
6.5S2	3677560	S-SNP	G->A	<i>lyxK</i>	
6.5S2	3715589	Deletion	1_bp	<i>ECB_03460</i>	
6.5S2	3762120	NS-SNP	C->A	<i>spoT</i>	A455D
6.5S2	3862401	S-SNP	C->T	<i>bglH</i>	
6.5S2	3866000	Deletion (Intergenic)	1_bp	<i>bglF</i> (98) / <i>bglG</i> (35)	
6.5S2	3866281	S-SNP	T->C	<i>bglG</i>	
6.5S2	3891539	Insertion	1_bp	<i>ESCRE3877</i>	
6.5S2	3894997	LargeDeletion	4900_bp	<i>rbsD</i> / <i>rbsK</i>	
6.5S2	3899899	IS_Insertion	IS150	<i>rbsK</i>	
6.5S2	3925659	NS-SNP	C->T	<i>gpp</i>	A125T
6.5S2	3935300	NS-SNP	A->G	<i>rffG</i>	T327A
6.5S2	3960103	NS-SNP	A->G	<i>uvrD</i>	H95R
6.5S2	3985793	NS-SNP	T->C	<i>ECB_03714</i>	T201A
6.5S2	4085858	Insertion (Intergenic)	1_bp	<i>cpxP</i> (93) / <i>fieF</i> (56)	
6.5S2	4085862	SNP (Intergenic)	T->T	<i>cpxP</i> (89) / <i>fieF</i> (60)	
6.5S2	4101025	NS-SNP	C->T	<i>hslU</i>	R69H
6.5S2	4117175	NS-SNP	A->G	<i>yijF</i>	V197A
6.5S2	4125279	NS-SNP	A->G	<i>pflD</i>	N393S
6.5S2	4127462	NS-SNP	G->A	<i>frwD</i>	R73Q
6.5S2	4137023	NS-SNP	A->G	<i>argH</i>	T105A
6.5S2	4164709	NS-SNP	A->G	<i>rpoB</i>	Y1285C
6.5S2	4185358	S-SNP	G->A	<i>purD</i>	
6.5S2	4202021	NS-SNP	C->T	<i>iclR</i>	R180H
6.5S2	4216015	NS-SNP	A->G	<i>yjbG</i>	S63G
6.5S2	4218174	Insertion	1_bp	<i>yjbH</i>	
6.5S2	4218862	Insertion	1_bp	<i>ESCRE4211</i>	
6.5S2	4225622	SNP (Intergenic)	G->A	<i>malE</i> (94) / <i>malK</i> (271)	
6.5S2	4235048	S-SNP	A->G	<i>plsB</i>	
6.5S2	4252196	NS-SNP	A->G	<i>uvrA</i>	V203A
6.5S2	4266390	SNP (Intergenic)	C->A	<i>acs</i> (307) / <i>nrfA</i> (86)	
6.5S2	4303399	NS-SNP	G->A	<i>phnC</i>	T194I
6.5S2	4324439	NS-SNP	A->G	<i>fumB</i>	F535L
6.5S2	4331214	Deletion (Intergenic)	1_bp	<i>yjdJ</i> (90) / <i>yjdK</i> (137)	
6.5S2	4334407	NS-SNP	A->G	<i>yjdL</i>	V242A
6.5S2	4361336	SNP (Intergenic)	A->G	<i>frdA</i> (18) / <i>poxA</i> (307)	

6.5S2	4361415	NS-SNP	C->T	<i>poxA</i>	A21V
6.5S2	4375768	Insertion	6_bp	<i>mutL</i>	
6.5S2	4403192	SNP (Intergenic)	C->T	<i>yjfY</i> (81) / <i>rpsF</i> (247)	
6.5S2	4455697	Deletion	1_bp	<i>pyrL</i>	
6.5S2	4471801	Insertion	1_bp	<i>yjgR</i>	
6.5S2	4484874	Insertion	1_bp	<i>IS4_insG</i>	
6.5S2	4506111	LargeDeletion	15,458_bp	<i>sgcR</i> / <i>fimB</i>	
6.5S2	4553692	NS-SNP	A->G	<i>ECB_04210</i>	E716G
6.5S2	4559960	NS-SNP	A->G	<i>hsdM</i>	L355P
6.5S2	4584759	NS-SNP	A->G	<i>yjiM</i>	F13L
6.5S2	4590035	NS-SNP	G->A	<i>dnaC</i>	R59C
6.5S2	4595558	Deletion	1_bp	<i>leuQ</i>	
6.5S2	4615700	NS-SNP	G->A	<i>nadR</i>	E58K
6.5S2	4628190	NS-SNP	G->A	<i>arcA</i>	L93F
6.5L4	92773	S-SNP	A->G	<i>mraZ</i>	
6.5L4	94365	NS-SNP	C->T	<i>ftsI</i>	P50L
6.5L4	95020	S-SNP	C->T	<i>ftsI</i>	
6.5L4	100745	S-SNP	A->G	<i>murD</i>	
6.5L4	115677	NS-SNP	A->G	<i>coaE</i>	V116A
6.5L4	129638	S-SNP	C->T	<i>aceF</i>	
6.5L4	178541	NS-SNP	T->C	<i>clcA</i>	L198S
6.5L4	205872	S-SNP	C->T	<i>lpxA</i>	
6.5L4	266234	S-SNP	A->G	<i>proA</i>	
6.5L4	294649	S-SNP	A->G	<i>ykgE</i>	
6.5L4	328102	NS-SNP	A->G	<i>codB</i>	T209A
6.5L4	332236	S-SNP	G->A	<i>cynS</i>	
6.5L4	379317	Insertion	1_bp	<i>mak</i>	
6.5L4	423643	NS-SNP	A->G	<i>ampG</i>	V203A
6.5L4	458996	NS-SNP	C->T	<i>kefA</i>	Q132*
6.5L4	474561	NS-SNP	A->G	<i>ybaL</i>	S249P
6.5L4	486377	NS-SNP	C->T	<i>cueR</i>	P73S
6.5L4	486436	S-SNP	A->G	<i>cueR</i>	
6.5L4	492859	S-SNP	C->T	<i>ybbP</i>	
6.5L4	501047	Insertion	1_bp	<i>ybbD</i>	
6.5L4	506968	NS-SNP	G->A	<i>gcl</i>	R296H
6.5L4	535466	NS-SNP	G->A	<i>sfmF</i>	C8Y
6.5L4	574371	NS-SNP	C->T	<i>cusS</i>	A435T
6.5L4	620503	Insertion	1_bp	<i>dsbG</i>	
6.5L4	624498	NS-SNP	A->G	<i>ybdR</i>	I120M
6.5L4	626325	NS-SNP	A->G	<i>rna</i>	S244P
6.5L4	649522	NS-SNP	C->T	<i>mrdA</i>	G427D
6.5L4	650661	NS-SNP	A->C	<i>mrdA</i>	F47L
6.5L4	664445	SNP	G->A	<i>gltL</i>	
6.5L4	681728	NS-SNP	T->C	<i>nagD</i>	S232G
6.5L4	787867	Insertion (Intergenic)	1_bp	<i>ybhC</i> (85) / <i>int</i> (50)	
6.5L4	804885	S-SNP	A->G	<i>bioC</i>	
6.5L4	806804	NS-SNP	A->G	<i>uvrB</i>	D218G
6.5L4	870659	SNP (Intergenic)	C->T	<i>yliG</i> (141) / <i>bssR</i> (72)	
6.5L4	889874	Deletion (Intergenic)	1_bp	<i>ECB_00829</i> (18) / <i>ECB_00830</i> (19)	
6.5L4	894055	NS-SNP	T->C	<i>EB3_00851</i>	Y232C

6.5L4	933517	S-SNP	T->C	<i>ybjD</i>	
6.5L4	997936	Insertion (Intergenic)	1_bp	<i>mukB</i> (203) / <i>ycbB</i> (58)	
6.5L4	999514	NS-SNP	A->G	<i>ycbB</i>	N459S
6.5L4	1025569	NS-SNP	C->T	<i>ycbY</i>	T212I
6.5L4	1030571	S-SNP	C->T	<i>pqiB</i>	
6.5L4	1047093	SNP (Intergenic)	C->T	<i>yccW</i> (62) / <i>yccX</i> (120)	
6.5L4	1053644	NS-SNP	A->G	<i>hyaE</i>	D67G
6.5L4	1057125	NS-SNP	G->A	<i>appB</i>	V247M
6.5L4	1124264	NS-SNP	G->A	<i>mdoG</i>	D121N
6.5L4	1161821	S-SNP	A->G	<i>yceD</i>	
6.5L4	1202311	S-SNP	A->G	<i>pepT</i>	
6.5L4	1234147	NS-SNP	G->A	<i>nhaB</i>	R241C
6.5L4	1246849	S-SNP	C->T	<i>treA</i>	
6.5L4	1301081	NS-SNP	A->G	<i>oppB</i>	M197V
6.5L4	1306561	SNP (Intergenic)	C->T	<i>ESCRE</i> 1299(49) / <i>kch</i> (6)	
6.5L4	1306976	S-SNP	A->G	<i>kch</i>	
6.5L4	1354410	NS-SNP	G->A	<i>sapA</i>	T357I
6.5L4	1357910	Insertion	1_bp	<i>puuA</i>	
6.5L4	1366183	NS-SNP	A->G	<i>pspF</i>	L33P
6.5L4	1371247	NS-SNP	A->G	<i>ycjN</i>	Q324R
6.5L4	1378694	Deletion	1_bp	<i>ycjU</i>	
6.5L4	1383106	S-SNP	G->A	<i>ycjX</i>	
6.5L4	1465797	NS-SNP	A->G	<i>ycdJ</i>	E134G
6.5L4	1510661	NS-SNP	C->T	<i>narV</i>	W36*
6.5L4	1609176	LargeDeletion	6292_bp	<i>ECB_01513</i> / <i>ECB_01523</i>	
6.5L4	1646650	S-SNP	A->G	<i>ynfL</i>	
6.5L4	1691513	Deletion	1_bp	<i>ydgR</i> (58) / <i>gst</i> (48)	
6.5L4	1711520	Insertion	1_bp	<i>ydhD</i> (109) / <i>ydhO</i> (225)	
6.5L4	1726247	NS-SNP	C->T	<i>ydhT</i>	A194T
6.5L4	1733754	NS-SNP	A->T	<i>pykF</i>	I264F
6.5L4	1787321	Insertion (Intergenic)	1_bp	<i>yniC</i> (155) / <i>ydjM</i> (8)	
6.5L4	1810725	S-SNP	T->C	<i>ydjX</i>	
6.5L4	1826622	S-SNP	T->C	<i>sppA</i>	
6.5L4	1881497	NS-SNP	T->C	<i>manX</i>	F291L
6.5L4	1928817	NS-SNP	T->C	<i>aspS</i>	D159G
6.5L4	1940852	Deletion	1_bp	<i>yecT</i>	
6.5L4	1944477	Deletion	1_bp	<i>flhB</i>	
6.5L4	1961078	NS-SNP	T->C	<i>araH</i>	K154E
6.5L4	2010137	NS-SNP	T->C	<i>yeeA</i>	K300E
6.5L4	2031432	LargeDeletion	23,293_bp	<i>cpsG</i> / <i>wcaJ</i>	
6.5L4	2078110	NS-SNP	T->C	<i>yegE</i>	F547L
6.5L4	2100286	LargeDeletion	22,146_bp	<i>ECB_misc_RNA_13</i> / <i>ECB_02013</i>	
6.5L4	2111272	Insertion	1_bp	<i>ECB_01999</i>	
6.5L4	2126793	S-SNP	T->C	<i>gatC</i>	
6.5L4	2132366	NS-SNP	C->T	<i>fbaB</i>	A110T
6.5L4	2140223	NS-SNP	C->T	<i>ECB_02034</i>	T60M
6.5L4	2142633	NS-SNP	G->A	<i>ESCRE</i> 2162	P61S
6.5L4	2171395	S-SNP	G->A	<i>bglX</i>	
6.5L4	2186143	S-SNP	T->C	<i>yeiT</i>	
6.5L4	2224713	NS-SNP	T->C	<i>yejA</i>	Y388H

6.5L4	2233001	Insertion	1_bp	<i>yejH</i>	
6.5L4	2246435	S-SNP	C->T	<i>ccmB</i>	
6.5L4	2303542	SNP (Intergenic)	A->G	<i>glpC(70) / yfaD(123)</i>	
6.5L4	2325601	S-SNP	A->G	<i>menD</i>	
6.5L4	2366466	S-SNP	G->A	<i>yfcI</i>	
6.5L4	2431702	NS-SNP	T->C	<i>yfdZ</i>	H273R
6.5L4	2544773	NS-SNP	C->T	<i>ppk</i>	P304L
6.5L4	2631815	NS-SNP	T->C	<i>ESCRE2622</i>	H4R
6.5L4	2633002	NS-SNP	C->T	<i>nadB</i>	R364C
6.5L4	2633980	NS-SNP	G->A	<i>yfiC</i>	P93L
6.5L4	2639803	S-SNP	G->A	<i>yfiF</i>	
6.5L4	2646652	S-SNP	T->C	<i>kgtP</i>	
6.5L4	2700659	NS-SNP	C->T	<i>proV</i>	Q365*
6.5L4	2704628	Insertion	1_bp	<i>ygaZ</i>	
6.5L4	2708050	S-SNP	T->C	<i>emrB</i>	
6.5L4	2735375	NS-SNP	T->C	<i>ascF</i>	M331T
6.5L4	2757641	NS-SNP	G->A	<i>ygbK</i>	M149I
6.5L4	2774203	Insertion (Intergenic)	1_bp	<i>IS186_insL2(13) / ESCRE2779(17)</i>	
6.5L4	2800862	NS-SNP	T->C	<i>relA</i>	D573G
6.5L4	2805234	NS-SNP	T->C	<i>barA</i>	V417A
6.5L4	2840130	NS-SNP	G->A	<i>ygdL</i>	P111L
6.5L4	2876176	NS-SNP	G->A	<i>araE</i>	R210C
6.5L4	2944404	SNP (Intergenic)	T->C	<i>serA(13) / rpiA(243)</i>	
6.5L4	2976544	S-SNP	T->C	<i>endA</i>	
6.5L4	2977749	NS-SNP	G->A	<i>gshB</i>	G41S
6.5L4	2997395	S-SNP	T->C	<i>ECB_02798</i>	
6.5L4	3002150	NS-SNP	C->T	<i>ECB_02800</i>	P863S
6.5L4	3022654	Deletion	2_bp	<i>ECB_02821</i>	
6.5L4	3037784	NS-SNP	G->A	<i>epsD</i>	T404M
6.5L4	3043945	Deletion	1_bp	<i>yghJ</i>	
6.5L4	3064150	NS-SNP	T->C	<i>ECB_02857</i>	N54S
6.5L4	3123048	NS-SNP	A->G	<i>yqiH</i>	Y111C
6.5L4	3145003	S-SNP	G->A	<i>dnaG</i>	
6.5L4	3147638	NS-SNP	C->T	<i>rpoD</i>	A592V
6.5L4	3150255	SNP (Intergenic)	G->A	<i>yqjI(118) / aer(36)</i>	
6.5L4	3155898	NS-SNP	T->C	<i>ebgA</i>	F150S
6.5L4	3170559	S-SNP	A->G	<i>ygiR</i>	
6.5L4	3189150	NS-SNP	T->C	<i>yhaM</i>	T179A
6.5L4	3203410	NS-SNP	T->C	<i>ECB_misc_RNA_46</i>	K18E
6.5L4	3203412	Insertion	1_bp	<i>ECB_misc_RNA_46</i>	
6.5L4	3218529	NS-SNP	G->A	<i>kbaY</i>	G52R
6.5L4	3220182	NS-SNP	A->G	<i>agaC</i>	T89A
6.5L4	3249780	NS-SNP	T->C	<i>infB</i>	Q490R
6.5L4	3250447	NS-SNP	G->A	<i>infB</i>	H268Y
6.5L4	3271434	NS-SNP	G->A	<i>murA</i>	P99L
6.5L4	3311431	S-SNP	T->C	<i>mdh</i>	
6.5L4	3373127	NS-SNP	T->C	<i>rpsE</i>	H3R
6.5L4	3385583	S-SNP	G->A	<i>gspD</i>	
6.5L4	3432156	NS-SNP	G->A	<i>friR</i>	R67H
6.5L4	3478428	S-SNP	G->A	<i>malQ</i>	

6.5L4	3481574	Insertion (Intergenic)	1_bp	<i>malP</i> (111) / <i>malT</i> (501)	
6.5L4	3610949	NS-SNP	T->C	<i>yhjJ</i>	S162G
6.5L4	3625452	Insertion	1_bp	<i>ECB_03382</i>	
6.5L4	3661344	SNP (Intergenic)	G->A	<i>xylA</i> (245) / <i>xylF</i> (121)	
6.5L4	3670123	SNP (Intergenic)	T->C	<i>malS</i> (38) / <i>avtA</i> (140)	
6.5L4	3677560	S-SNP	G->A	<i>lyxK</i>	
6.5L4	3711647	NS-SNP	T->C	<i>yibL</i>	V97A
6.5L4	3724621	NS-SNP	T->C	<i>secB</i>	T92A
6.5L4	3762120	NS-SNP	C->A	<i>spoT</i>	A455D
6.5L4	3775618	Insertion (Intergenic)	1_bp	<i>selC</i> (40) / <i>ECB_03516</i> (264)	
6.5L4	3866000	Deletion (Intergenic)	1_bp	<i>bglF</i> (98) / <i>bglG</i> (35)	
6.5L4	3891539	Insertion	2_bp	<i>ESCRE3877</i>	
6.5L4	3891917	Insertion	1_bp	<i>trkD</i>	
6.5L4	3894997	LargeDeletion	4900_bp	<i>rbsD</i> / <i>rbsK</i>	
6.5L4	3899899	IS_Insertion	IS150	<i>rbsK</i>	
6.5L4	4101025	NS-SNP	C->T	<i>hslU</i>	R69H
6.5L4	4111714	Deletion	1_bp	<i>metL</i>	
6.5L4	4111718	NS-SNP	A->C	<i>metL</i>	T593P
6.5L4	4176939	NS-SNP	T->C	<i>nudC</i>	L142P
6.5L4	4186693	NS-SNP	T->G	<i>purH</i>	K142T
6.5L4	4202021	NS-SNP	C->T	<i>iclR</i>	R180H
6.5L4	4218174	Insertion	1_bp	<i>yjbH</i>	
6.5L4	4218862	Insertion	1_bp	<i>ESCRE4211</i>	
6.5L4	4233773	S-SNP	G->A	<i>plsB</i>	
6.5L4	4272034	Insertion	1_bp	<i>nrfE</i>	
6.5L4	4291998	Insertion (Intergenic)	1_bp	<i>rpiR</i> (174) / <i>rpiB</i> (185)	
6.5L4	4303399	NS-SNP	G->A	<i>phnC</i>	T194I
6.5L4	4331214	Deletion (Intergenic)	1_bp	<i>yjdJ</i> (90) / <i>yjdK</i> (137)	
6.5L4	4332962	NS-SNP	C->T	<i>lysU</i>	G159D
6.5L4	4375768	Insertion	6_bp	<i>mutL</i>	
6.5L4	4387084	Insertion	1_bp	<i>rnr</i>	
6.5L4	4471093	S-SNP	T->C	<i>yjgQ</i>	
6.5L4	4484878	NS-SNP	A->G	<i>IS4_insG</i>	F149S
6.5L4	4504802	SNP (Intergenic)	G->A	<i>ESCRE4497</i> (555) / <i>IS1_insB</i> (15)	
6.5L4	4504883	Insertion (Intergenic)	1_bp	<i>ESCRE4497</i> (474) / <i>IS1_insB</i> (96)	
6.5L4	4530134	Deletion	1_bp	<i>fimH</i>	
6.5L4	4543192	NS-SNP	T->C	<i>yjiJ</i>	H133R
6.5L4	4560588	NS-SNP	C->T	<i>hsdM</i>	A146T
6.5L4	4584759	NS-SNP	A->G	<i>yjiM</i>	F13L
6.5L4	4595443	Deletion	1_bp	<i>leuP</i>	
6.5L4	4615700	NS-SNP	G->A	<i>nadR</i>	E58K
6.5L9	13825	S-SNP	C->T	<i>dnaK</i>	
6.5L9	92773	S-SNP	A->G	<i>mraZ</i>	
6.5L9	94365	NS-SNP	C->T	<i>ftsI</i>	P50L
6.5L9	100745	S-SNP	A->G	<i>murD</i>	
6.5L9	114648	NS-SNP	C->T	<i>yacG</i>	M1I
6.5L9	115677	NS-SNP	A->G	<i>coaE</i>	V116A
6.5L9	178541	NS-SNP	T->C	<i>clcA</i>	L198S
6.5L9	205872	S-SNP	C->T	<i>lpxA</i>	
6.5L9	255562	S-SNP	C->T	<i>dinB</i>	

6.5L9	266234	S-SNP	A->G	<i>proA</i>	
6.5L9	278231	Insertion	1_bp	<i>yagX</i>	
6.5L9	294649	S-SNP	A->G	<i>ykgE</i>	
6.5L9	306438	NS-SNP	C->T	<i>yahB</i>	W215*
6.5L9	324643	NS-SNP	T->C	<i>prpD</i>	V260A
6.5L9	332236	S-SNP	G->A	<i>cynS</i>	
6.5L9	339039	NS-SNP	G->A	<i>lacI</i>	A343V
6.5L9	352371	Deletion (Intergenic)	1_bp	<i>frmR</i> (53) / <i>yaiO</i> (133)	
6.5L9	357498	NS-SNP	C->T	<i>tauD</i>	R266W
6.5L9	388522	NS-SNP	C->T	<i>brnQ</i>	P170L
6.5L9	423643	NS-SNP	A->G	<i>ampG</i>	V203A
6.5L9	427311	SNP (Intergenic)	C->T	<i>tig</i> (71) / <i>clpP</i> (175)	
6.5L9	458996	NS-SNP	C->T	<i>kefA</i>	Q132*
6.5L9	459186	NS-SNP	C->T	<i>kefA</i>	A195V
6.5L9	465297	NS-SNP	C->T	<i>dnaX</i>	P380L
6.5L9	474561	NS-SNP	A->G	<i>ybaL</i>	S249P
6.5L9	486377	NS-SNP	C->T	<i>cueR</i>	P73S
6.5L9	501047	Insertion	1_bp	<i>ybbD</i>	
6.5L9	506968	NS-SNP	G->A	<i>gcl</i>	R296H
6.5L9	574371	NS-SNP	C->T	<i>cusS</i>	A435T
6.5L9	620503	Insertion	1_bp	<i>dsbG</i>	
6.5L9	626325	NS-SNP	A->G	<i>rna</i>	S244P
6.5L9	649522	NS-SNP	C->T	<i>mrda</i>	G427D
6.5L9	650661	NS-SNP	A->C	<i>mrda</i>	F47L
6.5L9	804885	S-SNP	A->G	<i>bioC</i>	
6.5L9	806804	NS-SNP	A->G	<i>uvrB</i>	D218G
6.5L9	823859	S-SNP	A->G	<i>rhIE</i>	
6.5L9	860968	Deletion	1_bp	<i>yliA</i>	
6.5L9	891665	Insertion	1_bp	<i>ECB_00831</i>	
6.5L9	894055	NS-SNP	T->C	<i>EB3_00851</i>	Y232C
6.5L9	911490	NS-SNP	G->A	<i>potF</i>	G334S
6.5L9	940430	S-SNP	C->T	<i>clpA</i>	
6.5L9	942704	Insertion	1_bp	<i>ESCRE0946</i>	
6.5L9	972525	NS-SNP	C->T	<i>ycaO</i>	W400*
6.5L9	996518	NS-SNP	T->C	<i>mukB</i>	L1034P
6.5L9	1007673	SNP (Intergenic)	A->G	<i>pncB</i> (40) / <i>pepN</i> (226)	
6.5L9	1023485	NS-SNP	C->T	<i>ycbW</i>	A148V
6.5L9	1030571	S-SNP	C->T	<i>pqiB</i>	
6.5L9	1038561	S-SNP	C->T	<i>yccR</i>	
6.5L9	1053644	NS-SNP	A->G	<i>hyaE</i>	D67G
6.5L9	1057125	NS-SNP	G->A	<i>appB</i>	V247M
6.5L9	1111957	SNP (Intergenic)	T->N	<i>IS3_insE</i> (176) / <i>serX</i> (226)	
6.5L9	1150423	S-SNP	C->T	<i>flgH</i>	
6.5L9	1165946	S-SNP	T->C	<i>fabG</i>	
6.5L9	1246849	S-SNP	C->T	<i>treA</i>	
6.5L9	1281732	S-SNP	C->T	<i>narG</i>	
6.5L9	1306561	SNP (Intergenic)	C->T	<i>ESCRE1299</i> (49) / <i>kch</i> (6)	
6.5L9	1328624	S-SNP	C->T	<i>sohB</i>	
6.5L9	1349560	Deletion (Intergenic)	1_bp	<i>fabI</i> (17) / <i>ESCRE1348</i> (152)	
6.5L9	1354410	NS-SNP	G->A	<i>sapA</i>	T357I

6.5L9	1371247	NS-SNP	A->G	<i>ycjN</i>	Q324R
6.5L9	1378694	Deletion	1_bp	<i>ycjU</i>	
6.5L9	1383106	S-SNP	G->A	<i>ycjX</i>	
6.5L9	1465797	NS-SNP	A->G	<i>ydcJ</i>	E134G
6.5L9	1512187	NS-SNP	A->G	<i>narY</i>	L273P
6.5L9	1543890	Insertion	1_bp	<i>gadC</i>	
6.5L9	1579819	NS-SNP	C->T	<i>yneH</i>	R159H
6.5L9	1589841	S-SNP	T->C	<i>ydeE</i>	
6.5L9	1609176	LargeDeletion	6292_bp	<i>ECB_01513 / ECB_01523</i>	
6.5L9	1646650	S-SNP	A->G	<i>ynfL</i>	
6.5L9	1657345	S-SNP	A->G	<i>ydgI</i>	
6.5L9	1673102	S-SNP	G->A	<i>uidA</i>	
6.5L9	1673642	SNP (Intergenic)	G->A	<i>uidA(15) / uidR(376)</i>	
6.5L9	1691513	Deletion (Intergenic)	1_bp	<i>ydgR(58) / gst(48)</i>	
6.5L9	1696554	S-SNP	A->G	<i>anmK</i>	
6.5L9	1711520	Insertion (Intergenic)	1_bp	<i>ydhD(109) / ydhO(225)</i>	
6.5L9	1733754	NS-SNP	A->T	<i>pykF</i>	I264F
6.5L9	1739609	S-SNP	G->A	<i>sufC</i>	
6.5L9	1758380	NS-SNP	G->A	<i>ydiR</i>	E238K
6.5L9	1763478	S-SNP	T->C	<i>pps</i>	
6.5L9	1826622	S-SNP	T->C	<i>sppA</i>	
6.5L9	1861828	S-SNP	T->C	<i>yeaV</i>	
6.5L9	1861915	Insertion	1_bp	<i>yeaV</i>	
6.5L9	1872182	NS-SNP	T->C	<i>yoaB</i>	W80R
6.5L9	1887642	NS-SNP	G->A	<i>yobH</i>	G47D
6.5L9	1900323	SNP (Intergenic)	T->C	<i>yebU(34) / yebV(84)</i>	
6.5L9	1906241	NS-SNP	T->C	<i>ptrB</i>	D392G
6.5L9	1928817	NS-SNP	T->C	<i>aspS</i>	D159G
6.5L9	1937934	Insertion	1_bp	<i>cutC</i>	
6.5L9	1940852	Deletion	1_bp	<i>yecT</i>	
6.5L9	1944477	Deletion	1_bp	<i>flhB</i>	
6.5L9	1944481	NS-SNP	A->C	<i>flhB</i>	F161V
6.5L9	2031432	LargeDeletion	23,293_bp	<i>cpsG / wcaJ</i>	
6.5L9	2032468	NS-SNP	C->N	<i>cpsG</i>	E110L
6.5L9	2102139	S-SNP	G->A	<i>ECB_01989</i>	
6.5L9	2111272	Insertion	1_bp	<i>ECB_01999</i>	
6.5L9	2120407	S-SNP	G->A	<i>ESCRE2129</i>	
6.5L9	2131459	Deletion (Intergenic)	1_bp	<i>gatY(182) / fbaB(118)</i>	
6.5L9	2132366	NS-SNP	C->T	<i>fbaB</i>	A110T
6.5L9	2140223	NS-SNP	C->T	<i>ECB_02034</i>	T60M
6.5L9	2142633	NS-SNP	G->A	<i>ESCRE2162</i>	P61S
6.5L9	2150450	NS-SNP	G->A	<i>ECB_02045</i>	R927Q
6.5L9	2171395	S-SNP	G->A	<i>bglX</i>	
6.5L9	2186143	S-SNP	T->C	<i>yeiT</i>	
6.5L9	2224713	NS-SNP	T->C	<i>yejA</i>	Y388H
6.5L9	2233005	NS-SNP	A->G	<i>yejH</i>	K396E
6.5L9	2236023	NS-SNP	A->G	<i>yejM</i>	T154A
6.5L9	2259446	S-SNP	T->C	<i>ada</i>	
6.5L9	2272055	S-SNP	T->C	<i>atoE</i>	
6.5L9	2273826	NS-SNP	G->A	<i>atoB</i>	V340I

6.5L9	2431702	NS-SNP	T->C	<i>yfdZ</i>	H273R
6.5L9	2544773	NS-SNP	C->T	<i>ppk</i>	P304L
6.5L9	2559557	S-SNP	T->C	<i>yfgM</i>	
6.5L9	2603920	NS-SNP	C->T	<i>yphH</i>	A120V
6.5L9	2625997	NS-SNP	T->C	<i>lepB</i>	Y269C
6.5L9	2631815	NS-SNP	T->C	<i>ESCRE2622</i>	H4R
6.5L9	2633002	NS-SNP	C->T	<i>nadB</i>	R364C
6.5L9	2633980	NS-SNP	G->A	<i>yfiC</i>	P93L
6.5L9	2639803	S-SNP	G->A	<i>yfiF</i>	
6.5L9	2645822	NS-SNP	G->A	<i>yfiM</i>	G51R
6.5L9	2646652	S-SNP	T->C	<i>kgtP</i>	
6.5L9	2679733	NS-SNP	T->C	<i>ECB_02510</i>	Y269H
6.5L9	2708050	S-SNP	T->C	<i>emrB</i>	
6.5L9	2734704	S-SNP	C->T	<i>ascF</i>	
6.5L9	2735375	NS-SNP	T->C	<i>ascF</i>	M331T
6.5L9	2748659	S-SNP	T->C	<i>hypE</i>	
6.5L9	2786400	SNP (Intergenic)	A->G	<i>ygcS(64) / ygcU(30)</i>	
6.5L9	2800862	NS-SNP	T->C	<i>relA</i>	D573G
6.5L9	2862418	NS-SNP	T->C	<i>ptsP</i>	T213A
6.5L9	2875072	NS-SNP	T->C	<i>ygeA</i>	D62G
6.5L9	2901574	SNP (Intergenic)	G->A	<i>ygfJ(240) / ygfK(81)</i>	
6.5L9	2908650	S-SNP	G->A	<i>xdhD</i>	
6.5L9	2916993	SNP (Intergenic)	T->C	<i>ygfT(128) / ygfU(308)</i>	
6.5L9	2939838	NS-SNP	T->C	<i>pepP</i>	M252V
6.5L9	2976544	S-SNP	T->C	<i>endA</i>	
6.5L9	2977749	NS-SNP	G->A	<i>gshB</i>	G41S
6.5L9	2997395	S-SNP	T->C	<i>ECB_02798</i>	
6.5L9	2997410	S-SNP	G->A	<i>ECB_02798</i>	
6.5L9	3005027	Insertion	1_bp	<i>ECB_02801</i>	
6.5L9	3037784	NS-SNP	G->A	<i>epsD</i>	T404M
6.5L9	3051100	S-SNP	G->A	<i>glcG</i>	
6.5L9	3064150	NS-SNP	T->C	<i>ECB_02857</i>	N54S
6.5L9	3096260	S-SNP	A->G	<i>sufI</i>	
6.5L9	3123048	NS-SNP	A->G	<i>yqiH</i>	Y111C
6.5L9	3145003	S-SNP	G->A	<i>dnaG</i>	
6.5L9	3147638	NS-SNP	C->T	<i>rpoD</i>	A592V
6.5L9	3151842	NS-SNP	T->C	<i>aer</i>	T18A
6.5L9	3178278	S-SNP	A->G	<i>exuT</i>	
6.5L9	3178797	S-SNP	A->G	<i>exuT</i>	
6.5L9	3189150	NS-SNP	T->C	<i>yhaM</i>	T179A
6.5L9	3192831	NS-SNP	C->T	<i>tdcF</i>	R105Q
6.5L9	3203410	NS-SNP	T->C	<i>ECB_misc_RNA_46</i>	K18E
6.5L9	3212290	NS-SNP	G->A	<i>kbaZ</i>	R150H
6.5L9	3218529	NS-SNP	G->A	<i>kbaY</i>	G52R
6.5L9	3220182	NS-SNP	A->G	<i>agaC</i>	T89A
6.5L9	3240042	Insertion	1_bp	<i>mtr</i>	
6.5L9	3242719	NS-SNP	G->A	<i>deaD</i>	P126S
6.5L9	3250447	NS-SNP	G->A	<i>infB</i>	H268Y
6.5L9	3268877	SNP (Intergenic)	T->C	<i>rplU(67) / ispB(192)</i>	
6.5L9	3277338	Insertion	1_bp	<i>kdsD</i>	

6.5L9	3339338	S-SNP	A->G	<i>fis</i>	
6.5L9	3373127	NS-SNP	T->C	<i>rpsE</i>	H3R
6.5L9	3383007	S-SNP	G->A	<i>gspA</i>	
6.5L9	3385583	S-SNP	G->A	<i>gspD</i>	
6.5L9	3389012	NS-SNP	G->A	<i>gspG</i>	V29M
6.5L9	3435550	NS-SNP	G->A	<i>yhfT</i>	Q157*
6.5L9	3471662	SNP (Intergenic)	T->C	<i>yhgG(170) / yhgA(33)</i>	
6.5L9	3478428	S-SNP	G->A	<i>malQ</i>	
6.5L9	3607090	S-SNP	T->C	<i>yhjG</i>	
6.5L9	3610949	NS-SNP	T->C	<i>yhjJ</i>	S162G
6.5L9	3670123	SNP (Intergenic)	T->C	<i>malS(38) / avtA(140)</i>	
6.5L9	3677560	S-SNP	G->A	<i>lyxK</i>	
6.5L9	3700901	NS-SNP	T->C	<i>rhsA</i>	V1342A
6.5L9	3762120	NS-SNP	C->A	<i>spoT</i>	A455D
6.5L9	3775618	Insertion (Intergenic)	1_bp	<i>selC(40) / ECB_03516(264)</i>	
6.5L9	3831736	NS-SNP	G->A	<i>dgoT</i>	A119V
6.5L9	3845799	S-SNP	T->C	<i>yidC</i>	
6.5L9	3849514	NS-SNP	C->T	<i>tnaA</i>	H142Y
6.5L9	3866000	Deletion (Intergenic)	1_bp	<i>bglF(98) / bglG(35)</i>	
6.5L9	3891539	Insertion	1_bp	<i>ESCRE3877</i>	
6.5L9	3891917	Insertion	1_bp	<i>trkD</i>	
6.5L9	3894997	LargeDeletion	4900_bp	<i>rbsD / rbsK</i>	
6.5L9	3899899	IS_Insertion	IS150	<i>rbsK</i>	
6.5L9	4038531	Deletion (Intergenic)	1_bp	<i>bipA(174) / yihL(43)</i>	
6.5L9	4057506	S-SNP	C->T	<i>yiiE</i>	
6.5L9	4061692	NS-SNP	A->G	<i>ECB_03779</i>	S760P
6.5L9	4101025	NS-SNP	C->T	<i>hslU</i>	R69H
6.5L9	4136247	NS-SNP	C->T	<i>argB</i>	A124V
6.5L9	4148218	S-SNP	C->T	<i>gltT</i>	
6.5L9	4176939	NS-SNP	T->C	<i>nudC</i>	L142P
6.5L9	4179792	Deletion (Intergenic)	1_bp	<i>yjaG(101) / hupA(86)</i>	
6.5L9	4202021	NS-SNP	C->T	<i>iclR</i>	R180H
6.5L9	4218174	Insertion	1_bp	<i>yjbH</i>	
6.5L9	4218862	Insertion	1_bp	<i>ESCRE4211</i>	
6.5L9	4233773	S-SNP	G->A	<i>plsB</i>	
6.5L9	4244918	Insertion	1_bp	<i>alr</i>	
6.5L9	4272034	Insertion	1_bp	<i>nrfE</i>	
6.5L9	4282687	NS-SNP	C->T	<i>yjcR</i>	R82H
6.5L9	4291998	Insertion (Intergenic)	1_bp	<i>rpiR(174) / rpiB(185)</i>	
6.5L9	4303399	NS-SNP	G->A	<i>phnC</i>	T194I
6.5L9	4331214	Deletion (Intergenic)	1_bp	<i>yjdJ(90) / yjdK(137)</i>	
6.5L9	4375768	Insertion	6_bp	<i>mutL</i>	
6.5L9	4387084	Insertion	1_bp	<i>rnr</i>	
6.5L9	4431036	S-SNP	T->C	<i>ytfN</i>	
6.5L9	4471093	S-SNP	T->C	<i>yjgQ</i>	
6.5L9	4501248	S-SNP	A->G	<i>fecA</i>	
6.5L9	4504802	SNP (Intergenic)	G->A	<i>ESCRE4497(555) / IS1_insB(15)</i>	
6.5L9	4504883	Insertion (Intergenic)	1_bp	<i>ESCRE4497(474) / IS1_insB(96)</i>	
6.5L9	4530134	Deletion	1_bp	<i>fimH</i>	
6.5L9	4584759	NS-SNP	A->G	<i>yjiM</i>	F13L

6.5L9	4615700	NS-SNP	G->A	<i>nadR</i>	E58K
-------	---------	--------	------	-------------	------

^a Position relative to the genome sequence of the REL606 ancestor strain²⁷.

^b SNP: Single Nucleotide Polymorphism; S-SNP: Synonymous-SNP; NS-SNP: Non-Synonymous-SNP.

^c For intergenic mutations, the two surrounding genes are indicated with the distance of each mutation in bp from either the last base of the stop codon or of the first base of the start codon of both adjacent genes. For large deletions, the two surrounding genes are indicated.

^d Amino-acid changes are given for NS-SNPs only. Stop codons are indicated by "*".

Supplementary Table 2. Mutations (68) shared by all four S and L clones.

Position ^a	Base Change	Type ^b	Protein Change ^c	Gene(s) ^d	Gene Function ^e	GO Category ^{ef}
94365	C->T	NS-SNP	P50L	<i>ftsI</i>	division-specific transpeptidase, penicillin-binding protein 3	GO:0009252, GO:0042493
115677	A->G	NS-SNP	V116A	<i>coaE</i>	dephospho-CoA kinase	ND
178541	T->C	NS-SNP	L198S	<i>clcA</i>	chloride channel, voltage-gated	ND
474561	A->G	NS-SNP	S249P	<i>ybaL</i>	putative transport protein	ND
486377	C->T	NS-SNP	P73S	<i>cueR</i>	copper responsive transcription regulator	GO:0006350
501047	1_bp	Insertion		<i>ybbD</i>	hypothetical protein	ND
506968	G->A	NS-SNP	R296H	<i>gcl</i>	glyoxylate carboligase	GO:0009441, GO:0009442
574371	C->T	NS-SNP	A435T	<i>cusS</i>	sensor kinase of copper sensing two-component system	GO:0006464
626325	A->G	NS-SNP	S244P	<i>rna</i>	RNase I	GO:0006401
649522	C->T	NS-SNP	G427D	<i>mrda</i>	peptidoglycan synthetase; penicillin-binding protein	GO:0009252, GO:0042493
806804	A->G	NS-SNP	D218G	<i>uvrB</i>	excinuclease ABC subunit B	GO:0006308, GO:0006281, GO:0009314, GO:0009432
1053644	A->G	NS-SNP	D67G	<i>hyaE</i>	processing of HyaA and HyaB proteins	GO:0009060, GO:0009061, GO:0006457
1057125	G->A	NS-SNP	V247M	<i>appB</i>	cytochrome oxidase bd-II, subunit II	ND
1354410	G->A	NS-SNP	T357I	<i>sapA</i>	homolog of <i>Salmonella</i> peptide transport periplasmic protein	GO:0009063
1465797	A->G	NS-SNP	E134G	<i>ycdJ</i>	hypothetical protein	ND
1711520	1_bp	Insertion (Intergenic)		<i>ydH</i> / <i>ydH</i> O	monothiol glutaredoxin / putative exported hydrolase	ND
1733754	A->T	NS-SNP	I264F	<i>pykF</i>	pyruvate kinase	GO:0006096, GO:0006113, GO:0009061
1940852	1_bp	Deletion		<i>yecT</i>	unknown CDS	ND
1944477	1_bp	Deletion		<i>flhB</i>	flagellar biosynthesis protein	ND
2111272	1_bp	Insertion		ECB_01999	conserved hypothetical protein	ND
2431702	T->C	NS-SNP	H273R	<i>yfdZ</i>	hypothetical protein	ND
2633002	C->T	NS-SNP	R364C	<i>nadB</i>	L-aspartate oxidase	GO:0009435
2633980	G->A	NS-SNP	P93L	<i>yfiC</i>	putative enzyme	ND
2735375	T->C	NS-SNP	M331T	<i>ascF</i>	PTS family enzyme IIBC component, cellobiose/salicin/arbutin-specific	GO:0016052
2800862	T->C	NS-SNP	D573G	<i>relA</i>	(p)ppGpp synthetase I	GO:0015949
2977749	G->A	NS-SNP	G41S	<i>gshB</i>	glutathione synthetase	GO:0006750
3037784	G->A	NS-SNP	T404M	<i>epsD</i>	general secretion pathway protein D precursor	ND
3064150	T->C	NS-SNP	N54S	ECB_02857	putative transcriptional regulator	ND
3123048	A->G	NS-SNP	Y111C	<i>yqiH</i>	putative periplasmic chaperone	GO:0006457
3203410	T->C	NS-SNP	K18E	ECB_misc_RNA_46	RNaseP_bact_a	ND

3218529	G->A	NS-SNP	G52R	<i>kbaY</i>	tagatose 6-phosphate aldolase 1, KbaY subunit	GO:0016052
3220182	A->G	NS-SNP	T89A	<i>agaC</i>	PTS family enzyme IIC component 1, N-acetylglactosamine-specific	ND
3250447	G->A	NS-SNP	H268Y	<i>infB</i>	translation initiation factor IF-2	GO:0006412
3610949	T->C	NS-SNP	S162G	<i>yhjJ</i>	putative peptidase	ND
3670123	T->C	SNP (Intergenic)		<i>malS / avtA</i>	alpha-amylase / valine-pyruvate aminotransferase	GO:0016052, GO:0000272, GO:0006508, GO:0006523
3762120	C->A	NS-SNP	A455D	<i>spoT</i>	(p)ppGpp synthetase II	GO:0015949, GO:0042594
3866000	1_bp	Deletion (Intergenic)		<i>bglF / bglG</i>	fused beta-glucoside-specific PTS enzymes / transcriptional antiterminator of the <i>bgl</i> operon	GO:0016052, GO:0006464, GO:0016052, GO:0006350
3891539	1_bp	Insertion		<i>ESCRE3877</i>	hypothetical protein	ND
3894997	4900_bp	Large Deletion		<i>rbsD / rbsK</i>	Multigenic	GO:0016052, GO:0016052, GO:0016052, GO:0016052, GO:0016052
3899899	IS150	IS_insertion		<i>rbsK</i>	ribokinase	GO:0016052
4101025	C->T	NS-SNP	R69H	<i>hslU</i>	ATP-dependent protease ATP-binding subunit	GO:0006508, GO:0006457, GO:0009266
4202021	C->T	NS-SNP	R180H	<i>iclR</i>	repressor of <i>aceBA</i> operon	GO:0006097, GO:0006350
4218174	1_bp	Insertion		<i>yjbH</i>	hypothetical protein	ND
4218862	1_bp	Insertion		<i>ESCRE4211</i>	conserved hypothetical protein	ND
4303399	G->A	NS-SNP	T194I	<i>phnC</i>	ATP-binding component of phosphonate transport	GO:0006793
4331214	1_bp	Deletion (Intergenic)		<i>yjdJ / yjdK</i>	putative acyltransferase / conserved hypothetical protein	ND
4375768	6_bp	Insertion		<i>mutL</i>	DNA mismatch repair protein	GO:0006281
4584759	A->G	NS-SNP	F13L	<i>yjiM</i>	putative transcriptional regulator	GO:0006355
4615700	G->A	NS-SNP	E58K	<i>nadR</i>	nicotinamide-nucleotide adenyltransferase	GO:0009435, GO:0006350
100745	A->G	S-SNP		<i>murD</i>		
205872	C->T	S-SNP		<i>lpxA</i>		
266234	A->G	S-SNP		<i>proA</i>		
294649	A->G	S-SNP		<i>ykgE</i>		
332236	G->A	S-SNP		<i>cynS</i>		
804885	A->G	S-SNP		<i>bioC</i>		
1030571	C->T	S-SNP		<i>pqiB</i>		
1246849	C->T	S-SNP		<i>treA</i>		
1383106	G->A	S-SNP		<i>ycjX</i>		
1646650	A->G	S-SNP		<i>ynfL</i>		
1826622	T->C	S-SNP		<i>sppA</i>		
2171395	G->A	S-SNP		<i>bglX</i>		
2186143	T->C	S-SNP		<i>yeiT</i>		
2639803	G->A	S-SNP		<i>yfiF</i>		

2646652	T->C	S-SNP	<i>kgtP</i>
2976544	T->C	S-SNP	<i>endA</i>
3385583	G->A	S-SNP	<i>gspD</i>
3478428	G->A	S-SNP	<i>malQ</i>
3677560	G->A	S-SNP	<i>lyxK</i>

^a Position relative to the genome sequence of the REL606 ancestor strain²⁷.

^b SNP: Single Nucleotide Polymorphism; S-SNP: Synonymous-SNP; NS-SNP: Non-Synonymous-SNP.

^c Amino-acid changes are given for NS-SNPs only.

^d For intergenic mutations and large deletions, the two surrounding genes are indicated.

^e Informations are not given for S-SNPs.

^f Cellular functions correspond to the gene ontology (GO) cellular process categories³⁵. ND: Not Defined.

Supplementary Table 3. Mutations (55) shared by S clones only.

Position ^a	Base Change	Type ^b	Protein Change ^c	Gene(s) ^d	Gene Function ^e	GO Category ^{ef}
12074	T->A	SNP (Intergenic)		<i>yaal / dnaK</i>	conserved hypothetical protein / chaperone Hsp70	GO:0006457, GO:0006970
269551	C->T	NS-SNP	C523Y	<i>yagR</i>	putative molybdenum cofactor-binding oxidoreductase	ND
451466	C->T	NS-SNP	G174S	<i>maa</i>	maltose O-acetyltransferase	GO:0016052
794715	2_bp	Insertion		<i>ECB_00735</i>	fragment of conserved hypothetical protein	ND
825047	A->G	SNP (Intergenic)		<i>rhIE / ybiA</i>	RNA helicase / conserved hypothetical protein	GO:0009451
845838	1_bp	Insertion		<i>mntR</i>	DNA-binding transcriptional regulator of <i>mntH</i>	ND
950771	A->G	NS-SNP	T164A	<i>ftsK</i>	cell division protein	GO:0009432
1159018	T->C	SNP (Intergenic)		<i>rne / yceQ</i>	endoribonuclease / conserved hypothetical protein	GO:0006401
1614900	IS3	IS_insertion		<i>ECB_01523</i>	fragment of conserved hypothetical protein from bacteriophage origin	ND
1689708	1_bp	Insertion (Intergenic)		<i>nth / ydgR</i>	DNA glycosylase and apyrimidinic lyase / putative oligopeptide transporter	GO:0006308, GO:0009314
1815710	T->C	NS-SNP	V409A	<i>ynjC</i>	putative transport protein (ABC superfamily, membrane)	ND
1969213	1_bp	Deletion		<i>yecA</i>	putative protein translocase subunit	ND
2094204	T->C	NS-SNP	L904P	<i>mdtC</i>	multidrug efflux system, subunit C	ND
2175382	C->T	NS-SNP	V243M	<i>pbpG</i>	D-alanyl-D-alanine endopeptidase; penicillin-binding proteins 7 and 8	GO:0009252, GO:0042493
2213468	T->C	NS-SNP	N29S	<i>fruK</i>	1-phosphofructokinase	GO:0006096
2216363	G->A	NS-SNP	R41C	<i>yeiW</i>	conserved protein	ND
2378672	T->C	NS-SNP	H186R	<i>truA</i>	tRNA pseudouridine synthase A	GO:0006418
2418405	G->A	NS-SNP	D142N	<i>evgA</i>	putative positive transcription regulator	GO:0006350
2442609	A->G	SNP (Intergenic)		<i>ypdH / glk</i>	putative enzyme IIB component of PTS / glucokinase	GO:0009401, GO:0016052
2522405	T->C	NS-SNP	V126A	<i>hyfA</i>	hydrogenase 4 Fe-S subunit	ND
2798243	G->A	NS-SNP	H118Y	<i>pyrG</i>	CTP synthetase	GO:0015949
2897437	G->A	NS-SNP	M74I	<i>yqeA</i>	carbamate kinase	ND
2899613	IS150	IS_insertion		<i>yqeB</i>	putative synthase	ND
3067566	T->C	NS-SNP	K160R	<i>yghS</i>	hypothetical protein	ND
3158462	G->A	NS-SNP	A1005T	<i>ebgA</i>	beta-D-galactosidase, <i>ebg</i> operon, alpha subunit, cryptic gene	GO:0016052
3162869	G->A	NS-SNP	V389I	<i>ygjK</i>	putative enzyme	ND
3168389	C->T	NS-SNP	D102N	<i>ygjO</i>	putative enzyme	ND
3220975	G->A	NS-SNP	D89N	<i>agaD</i>	PTS family enzyme IID, N-acetylglucosamine-specific	GO:0046349
3256388	T->C	NS-SNP	S154G	<i>yhbX</i>	putative alkaline phosphatase I	ND
3419452	C->T	NS-SNP	R25H	<i>yhfG</i>	hypothetical protein	ND

3425074	T->C	NS-SNP	V56A	<i>nirC</i>	nitrite reductase activity	GO:0006807
3509731	T->C	NS-SNP	D150G	<i>gntR</i>	regulator of gluconate (<i>gnt</i>) operon	GO:0009255, GO:0006350
3512018	A->T	NS-SNP	Y81N	<i>yhhX</i>	putative NAD(P)-binding dehydrogenase	ND
3514265	T->C	NS-SNP	V317A	<i>yhhZ</i>	hypothetical protein	ND
3925659	C->T	NS-SNP	A125T	<i>gpp</i>	guanosine pentaphosphatase, also has exopolyphosphatase activity	GO:0015949, GO:0006793, GO:0042594
3985793	T->C	NS-SNP	T201A	<i>ECB_03714</i>	putative enzyme, putative Succinyl-CoA synthetase, alpha subunit	ND
4085858	1_bp	Insertion (Intergenic)		<i>cpxP / fieF</i>	periplasmic protein combats stress / zinc transporter	GO:0006350, GO:0006805
4117175	A->G	NS-SNP	V197A	<i>yijF</i>	hypothetical protein	ND
4266390	C->A	SNP (Intergenic)		<i>acs / nrfA</i>	acetyl-CoA synthetase / formate-dependent nitrite reductase	GO:0016052, GO:0045733, GO:0009061, GO:0017004
4334407	A->G	NS-SNP	V242A	<i>yjdL</i>	putative di-/tripeptide transport protein (POT family)	ND
4361415	C->T	NS-SNP	A21V	<i>poxA</i>	lysyl-tRNA synthetase	GO:0006418
4471801	1_bp	Insertion		<i>yjgR</i>	putative enzyme contains P-loop	ND
4506111	15,458_bp	Large Deletion		<i>sgcR / fimB</i>	Multigenic	GO:0006350, GO:0009401, GO:0016070, GO:0009297, GO:0006310, GO:0006261
4590035	G->A	NS-SNP	R59C	<i>dnaC</i>	DNA replication protein	GO:0006261
4628190	G->A	NS-SNP	L93F	<i>arcA</i>	negative response regulator of genes in aerobic pathways	GO:0006350
332926	C->T	S-SNP		<i>cynX</i>		
1149462	G->A	S-SNP		<i>flgG</i>		
2169035	C->T	S-SNP		<i>yehY</i>		
2452949	G->A	S-SNP		<i>gltX</i>		
2784199	A->G	S-SNP		<i>ygcQ</i>		
3075592	G->A	S-SNP		<i>hybC</i>		
3343907	C->T	S-SNP		<i>acrF</i>		
3566519	A->G	S-SNP		<i>yhiM</i>		
4185358	G->A	S-SNP		<i>purD</i>		
4235048	A->G	S-SNP		<i>plsB</i>		

^a Position relative to the genome sequence of the REL606 ancestor strain²⁷.

^b SNP: Single Nucleotide Polymorphism; S-SNP: Synonymous-SNP; NS-SNP: Non-Synonymous-SNP.

^c Amino-acid changes are given for NS-SNPs only.

^d For intergenic mutations and large deletions, the two surrounding genes are indicated.

^e Informations are not given for S-SNPs.

^f Cellular functions correspond to the gene ontology (GO) cellular process categories³⁵. ND: Not Defined.

Supplementary Table 4. Mutations (36) shared by L clones only.

Position ^a	Base Change	Type ^b	Protein Change ^c	Gene(s) ^d	Gene Function ^e	GO Category ^{ef}
423643	A->G	NS-SNP	V203A	<i>ampG</i>	muropeptide transport protein (MFS family)	GO:0000270
458996	C->T	NS-SNP	Q132*	<i>kefA</i>	component of the MscS mechanosensitive channel	ND
620503	1_bp	Insertion		<i>dsbG</i>	disulfide isomerase, thiol-disulphide oxidase, periplasmic	GO:0006457
650661	A->C	NS-SNP	F47L	<i>mrda</i>	peptidoglycan synthetase; penicillin-binding protein	GO:0009252, GO:0042493
894055	T->C	NS-SNP	Y232C	<i>mukB</i>	condesin subunit B	ND
1306561	C->T	SNP (Intergenic)		<i>ESCRE1299 / kch</i>	conserved hypothetical protein / voltage-gated potassium channel	ND
1371247	A->G	NS-SNP	Q324R	<i>ycjN</i>	putative ABC transporter periplasmic binding protein	ND
1378694	1_bp	Deletion		<i>ycjU</i>	putative beta phosphoglucomutase	ND
1609176	6292_bp	Large Deletion		<i>ECB_01513 / ECB_01523</i>	Multigenic	ND
1691513	1_bp	Deletion (Intergenic)		<i>ydgR / gst</i>	putative oligopeptide transporter / glutathione S-transferase	ND
1928817	T->C	NS-SNP	D159G	<i>aspS</i>	aspartyl-tRNA synthetase	GO:0006418
2031432	23,293_bp	Large Deletion		<i>cpsG / wcaJ</i>	Multigenic	GO:0009226, GO:0009242, GO:0009269, GO:0009243, GO:0009226, GO:0009243, GO:0009103, GO:0019305, GO:0009243, GO:0019305, GO:0009103, GO:0009243, GO:0009226, GO:0009243, GO:0009226, GO:0009435, GO:0009226, GO:0009242, GO:0009242, GO:0009242, GO:0009242
2132366	C->T	NS-SNP	A110T	<i>fbaB</i>	fructose-bisphosphate aldolase	GO:0006096
2140223	C->T	NS-SNP	T60M	<i>ECB_02034</i>	conserved hypothetical protein	ND
2142633	G->A	NS-SNP	P61S	<i>ESCRE2162</i>	hypothetical protein	ND
2224713	T->C	NS-SNP	Y388H	<i>yejA</i>	putative oligopeptide transport protein (ABC superfamily)	ND
2544773	C->T	NS-SNP	P304L	<i>ppk</i>	polyphosphate kinase	GO:0006793
2631815	T->C	NS-SNP	H4R	<i>ESCRE2622</i>	conserved hypothetical protein	ND
3147638	C->T	NS-SNP	A592V	<i>rpoD</i>	RNA polymerase sigma factor	GO:0006350

3189150	T->C	NS-SNP	T179A	<i>yhaM</i>	hypothetical protein	ND
3373127	T->C	NS-SNP	H3R	<i>rpsE</i>	30S ribosomal protein S5	GO:0006412
3775618	1_bp	Insertion (Intergenic)		<i>selC</i> / <i>ECB_03516</i>	tRNA-Sec / Putative prophage integrase	ND
3891917	1_bp	Insertion		<i>trkD</i>	potassium transporter	GO:0006970
4176939	T->C	NS-SNP	L142P	<i>nudC</i>	NADH pyrophosphatase	ND
4272034	1_bp	Insertion		<i>nrfE</i>	formate-dependent nitrite reductase	GO:0009061, GO:0017004, GO:0006457
4291998	1_bp	Insertion (Intergenic)		<i>rpiR</i> / <i>rpiB</i>	DNA-binding transcriptional repressor / ribose 5-phosphate isomerase	GO:0016052, GO:0009052, GO:0006350, GO:0016052, GO:0009052
4387084	1_bp	Insertion		<i>rnr</i>	RNase R	GO:0006401,
4504802	G->A	SNP (Intergenic)		<i>ESCRE4497</i> / <i>IS1_insB</i>	hypothetical protein / IS1 transposase	ND
4504883	1_bp	Insertion (Intergenic)		<i>ESCRE4497</i> / <i>IS1_insB</i>	hypothetical protein / IS1 transposase	ND
4530134	1_bp	Deletion		<i>fimH</i>	minor fimbrial subunit	GO:0009101
92773	A->G	S-SNP		<i>mraZ</i>		
2708050	T->C	S-SNP		<i>emrB</i>		
2997395	T->C	S-SNP		<i>ECB_02798</i>		
3145003	G->A	S-SNP		<i>dnaG</i>		
4233773	G->A	S-SNP		<i>plsB</i>		
4471093	T->C	S-SNP		<i>yjgQ</i>		

^a Position relative to the genome sequence of the REL606 ancestor strain²⁷.

^b SNP: Single Nucleotide Polymorphism; S-SNP: Synonymous-SNP; NS-SNP: Non-Synonymous-SNP.

^c Amino-acid changes are given for NS-SNPs only.

^d For intergenic mutations and large deletions, the two surrounding genes are indicated.

^e Informations are not given for S-SNPs.

^f Cellular functions correspond to the gene ontology (GO) cellular process categories³⁵. ND: Not Defined.

Supplementary Table 5. Functional GO categories significantly affected by mutations in evolved S and L clones.

	P Value (Fisher exact test)	GO Category	Function
Mutations shared by S and L	0.033262969495	GO:0009401	phosphoenolpyruvate-dependent sugar phosphotransferase system
Mutations shared by S only	0.020858137938908	GO:0009401	phosphoenolpyruvate-dependent sugar phosphotransferase system
Mutations shared by L only	0.035222580449699	GO:0006351	transcription, DNA-dependent
	0.00038381824711536	GO:0009226	nucleotide-sugar biosynthetic process
	0.0038919024593508	GO:0009242	colanic acid biosynthetic process
	0.00015815395606571	GO:0009243	O-antigen biosynthetic process
	0.034470910820018	GO:0016052	carbohydrate catabolic process
	0.0064795335821193	GO:0019305	dTDP-rhamnose biosynthetic process

Three sets of mutations were analyzed: the 68 shared by all four S and L genomes

(Supplementary Table 2), the 55 shared by S genomes only (Supplementary Table 3), and the 36 shared by L genomes only (Supplementary Table 4). The analyses were performed using all mutations except the synonymous ones. Fisher exact tests were performed for each GO category³⁵ in each of the three datasets by comparing the proportion of mutated genes in each GO category to the proportion of mutated genes in the genome (<http://marne.u707.jussieu.fr/biostatgv/?module=tests/fisher>). Shown here are only the GO categories that are significantly affected by mutations in each of the three datasets. A significant proportion of mutations in the two first datasets (pre-divergence and S only) targeted phosphotransferase transport systems, while L-specific mutations were more broadly distributed among GO categories.

Supplementary Table 6. Non-synonymous to synonymous rate ratios (Ka/Ks).

	Ka	Ks	omega
Mutations shared by S and L	1.21×10^{-5}	1.41×10^{-5}	0.8578
Mutations shared by S only	1.14×10^{-5}	7.41×10^{-6}	1.539
Mutations shared by L only	5.67×10^{-6}	5.56×10^{-6}	1.02

For each set of mutations, that affected coding regions, along each evolutionary branch, Ka and

Ks were computed using Comeron's method³⁶, implemented in libseq³⁷. The ancestral genome

was used to compute the number of potential synonymous and non-synonymous mutations.

Supplementary Table 7. Changes in SpoT, ArcA, ArcB and GntR in the twelve evolving populations of the long-term experiment.

Population	SpoT ³⁸	ArcA ^a	ArcB ^a	GntR ^a
Ara-1	K662I		R57C	
Ara-2 (S lineage)	A455D	L93F		D150G
Ara-3			V70L, Q79L	
Ara-4	Y389C		K175E, G441S, N658S	
Ara-5				
Ara-6	K607T	ΔP216M217R218		
Ara+1		T3P		
Ara+2	R575L	D29A		
Ara+3	A189V	E213K		
Ara+4	R209H	I42F		
Ara+5		C233F		
Ara+6	N454I, I602L		K343Q	K19T

All changes are given for the proteins. In this study, we focused on the changes in SpoT from Ara-1, in ArcA from Ara+1 and Ara+2, and in ArcB and GntR from Ara+6 (shown in bold) and used them to construct isogenic strains except for these evolved alleles. Two strains were derived from the ancestor carrying the three S evolved *spoT*, *arcA* and *gntR* alleles, one with the Ara-1 evolved *spoT* allele and one with the Ara+1 evolved *arcA* allele substituting the corresponding S alleles, respectively. Four strains were derived from 6.5S1: one in which the *gntR* allele was replaced by its Ara+6 counterpart, two with the *arcA* allele replaced by each Ara+1 and Ara+2 allele, and one with the ancestral *arcA* allele and the Ara+6 evolved *arcB* allele.

^a Unpublished data.

PARTIE II - Résultats supplémentaires

I. Qu'en est-il de la lignée L ?

Les trois mutations *spoT*, *arcA* et *gntR* sont impliquées dans l'événement de divergence, en conférant les phénotypes caractéristiques de la lignée S. Cependant, aucune information n'est disponible pour la lignée L et, à l'extrême, la question pourrait même se poser de savoir dans quelle mesure le génotype L *stricto sensu* est un pré-requis pour l'établissement du polymorphisme. Nous avons réalisé des expériences préliminaires afin d'étudier cette question.

Des expériences d'invasion réciproque ont été réalisées pour mesurer la capacité d'invasion de chacun des deux clones suivants : le clone évolué 6,5S1 (clone S isolé à 6500 générations) et le clone ancestral portant les allèles évolués des trois gènes *spoT*, *arcA*, et *gntR*. Cependant, les expériences d'invasion réciproque n'ont cette fois pas été réalisées contre un clone évolué de la lignée L comme précédemment, mais avec l'ancêtre comme compétiteur. Pour les deux clones, des interactions dépendantes de la fréquence ont été observées, c'est-à-dire que la valeur sélective de chacun d'entre eux était supérieure à celui de l'ancêtre quand leur fréquence initiale était faible (Figure 33a). Cependant, aucune invasion réciproque n'a été observée et l'ancêtre, quand il est rare, n'est capable d'envahir ni le clone 6,5S1, ni le clone ancestral portant les trois mutations (Figure 33a). Cette incapacité reflète probablement la valeur sélective meilleure de ces deux derniers clones par rapport à l'ancêtre (Figure supplémentaire 3b de l'article joint).

Pour éviter cette différence importante de valeur sélective relatif entre les compétiteurs, les expériences d'invasion réciproque ont été répétées (Figure 33b), mais cette fois entre le clone ancestral portant les allèles évolués des trois gènes *spoT*, *arcA*, et *gntR*, et le clone ancestral portant la mutation *spoT* qui présente une valeur sélective meilleure que l'ancêtre (Figure supplémentaire 3b de l'article joint). Cette fois, des interactions dépendantes de la fréquence avec invasion réciproque des compétiteurs ont bien été observées (Figure 33b), comme cela se produit entre les lignées S et L. Ces résultats montrent que les trois mutations *spoT*, *arcA* et *gntR* confèrent des capacités d'invasion réciproque non seulement par rapport à des clones évolués L, mais également par rapport à l'ancêtre portant la mutation *spoT*, un type de clones présent dans la population Ara-2 bien avant l'événement de divergence (la mutation *spoT* est en effet substituée dans la population Ara-2 dès 500 générations, alors que les clones S ne sont détectés qu'entre 6000 et 6500 générations). Ces résultats pourraient ainsi suggérer que l'émergence de la lignée L pourrait ne pas forcément être un pré-requis pour l'établissement du polymorphisme, ou que la mutation *spoT* pourrait à elle seule être responsable du phénotype correspondant à un clone L.

Les clones S étant capables de se développer dans un surnageant provenant de cultures de clones L, nous avons alors testé la capacité du clone ancêtre portant les trois mutations *spoT*, *arcA* et *gntR* à croître dans le surnageant provenant d'une culture du clone ancêtre portant la mutation *spoT*. Aucune croissance n'a été observée (résultats non montrés). Bien que le triple mutant soit capable d'envahir le clone ancêtre portant la mutation *spoT*, il s'agit donc probablement d'un phénomène différent de celui permettant les interactions entre les lignées S et L (voir plus bas). Nous avons donc

initié les analyses permettant d'identifier les mutations à la base de l'origine de la lignée L.

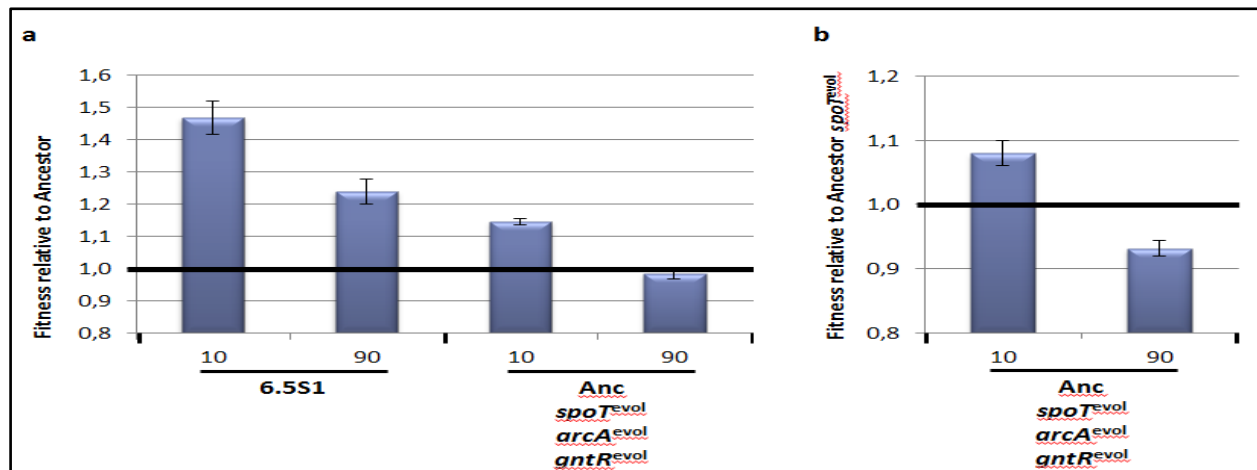


Figure 33 : Expériences d'invasion réciproque entre d'une part des clones portant les mutations des gènes *spoT*, *arcA* et *gntR* et d'autre part l'ancêtre (a) ou un dérivé de l'ancêtre portant l'allèle évolué *spoT* (b). (a) Valeur sélective relative à l'ancêtre après des expériences d'invasion réciproque avec, de gauche à droite, le clone évolué 6,5S1 et l'ancêtre portant les allèles évolués *spoT*, *arcA*, et *gntR*. Pour distinguer les compétiteurs, un révertant Ara⁺ du clone ancestral Ara⁻ a été utilisé dans toutes les expériences d'invasion. **(b)** Valeur sélective relative à l'ancêtre portant l'allèle évolué *spoT* après des expériences d'invasion réciproque avec l'ancêtre portant les allèles évolués *spoT*, *arcA* et *gntR*. Pour distinguer les compétiteurs, un révertant Ara⁺ de l'ancêtre portant l'allèle évolué *spoT* a été utilisé dans les expériences d'invasion. Les détails des expériences d'invasion et des calculs de valeur sélective sont identiques à celles données dans l'article joint (Pluain et al).

II. Histoire évolutive de la lignée L

Afin de déterminer les bases génétiques conduisant à la lignée L, nous avons réalisé des expériences d'invasion réciproque entre d'une part chacun des clones 6,5S1 et 6K3 (isolé avant l'émergence du polymorphisme, qui ne possède pas les phénotypes caractéristiques de la lignée S, mais qui porte des marqueurs génétiques spécifiques de la lignée S ; voir article), et d'autre part chacun des clones 6,5L4 (clone L isolé à 6500

générations) et 5000.1, 5000.5, 6000.2 et 6000.6 isolés respectivement à 5000 et 6000 générations. Ces quatre derniers clones portent un marqueur génétique diagnostique des clones L, en l'occurrence l'insertion spécifique d'une copie d'IS3 (Rozen et al 2005). Ces quatre clones sont donc présents avant l'émergence du polymorphisme, et se situent sur la ligne de descendance de la lignée L. Dans tous les cas, les deux clones isolés à 5000 générations et les deux clones isolés à 6000 générations ont conduit à des résultats similaires et seuls ceux obtenus avec les clones 5000.5 et 6000.2 sont présentés ici. Les expériences d'invasion ont bien confirmé que les clones évolués contemporains 6,5S1 et 6,5L4 étaient capables d'invasion réciproque (Figure 34). En revanche, bien que les clones 5000.5 et 6000.2 soient capables d'envahir 6,5S1, la réciproque n'est pas vraie (Figure 34). Une situation similaire a été observée en utilisant le clone 6K3, qui se situe sur la ligne de descendance de la lignée S, comme compétiteur (Figure 35). Alors que tous les clones 5000.5, 6000.2 et 6,5L4 sont capables de l'envahir, lui n'est capable d'envahir aucun de ces trois clones, confirmant également les données présentées dans l'article (voir article joint). En effet, nous avons déjà démontré qu'il existait des différences génétiques entre les clones 6K3 et 6,5S1 qui ne permettaient pas au clone 6K3 de posséder tous les phénotypes de clones S (en l'occurrence la mutation *gntR* est absente chez 6K3, voir l'article joint). Ces résultats suggèrent donc que des mutations importantes pour la capacité d'invasion des clones L sont déjà présentes dans les clones évolués isolés à 5000 générations. En revanche, le génotype complet conduisant à la lignée L n'est pas encore présent, puisque la lignée S ne peut pas envahir ces clones.

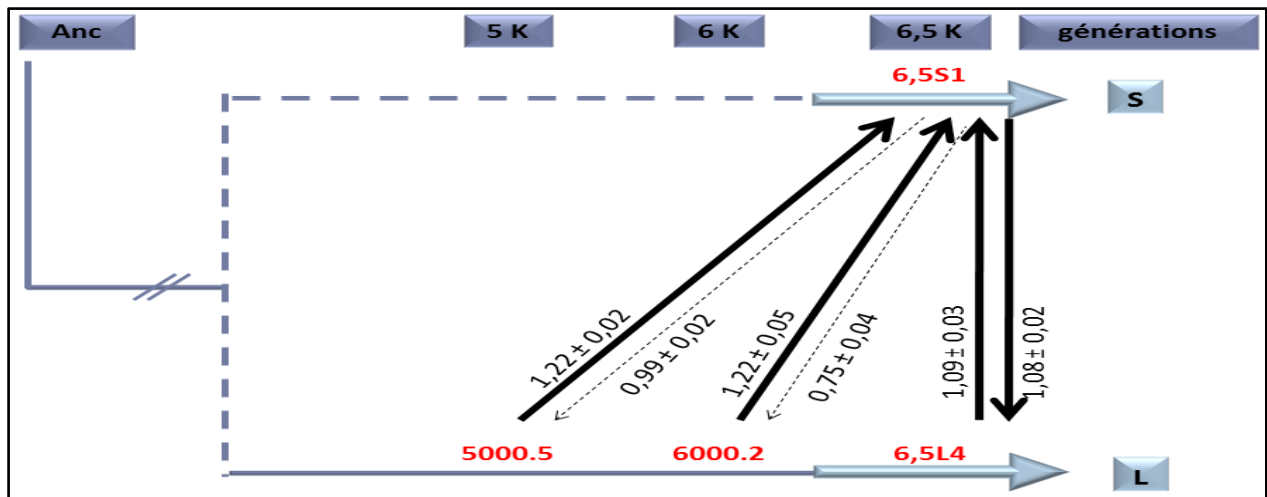


Figure 34 : Expériences d'invasion réciproque contre 6,5S1. Des expériences d'invasion réciproque ont été effectuées entre le clone 6,5S1 et chacun des clones 5000.5, 6000.2 et 6,5L4. Les flèches pleines et hachurées désignent, respectivement, la capacité ou non d'invasion du compétiteur, la base des flèches partant toujours du compétiteur initialement rare. Les détails des expériences d'invasion et des calculs de valeur sélectiv sont identiques à celles données dans l'article joint (Pluain et al). Les deux compétiteurs ont été distingués sur la base de la capacité ou non à utiliser le maltose comme source de carbone. En effet, le clone 6,5S1 est Mal⁻, alors que les autres sont Mal⁺.

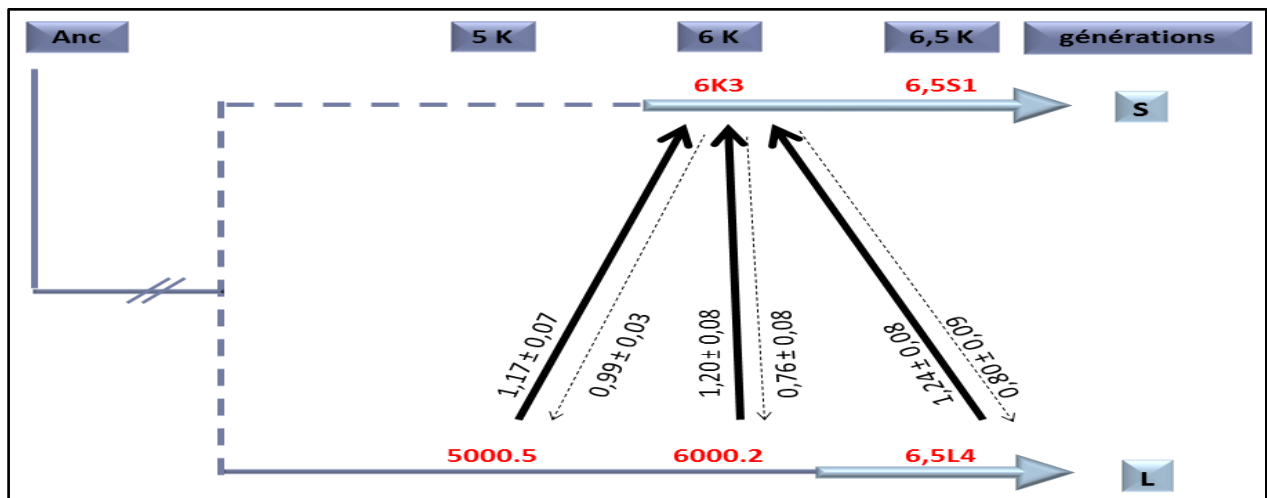


Figure 35 : Expériences d'invasion réciproque contre 6K3. Des expériences d'invasion réciproques ont été effectuées entre le clone 6K3 et chacun des clones 5000.5, 6000.2 et 6,5L4. Les flèches pleines et hachurées désignent, respectivement, la capacité ou non d'invasion du compétiteur, la base des flèches partant toujours du compétiteur initialement rare. Les détails des expériences d'invasion et des calculs de valeur sélective sont identiques

à celles données dans l'article joint (Plucaïn et al). Les deux compétiteurs ont été distingués sur la base de la capacité ou non à utiliser le maltose comme source de carbone. En effet, le clone 6K3 est Mal⁻, alors que les autres sont Mal⁺.

Une des hypothèses pouvant expliquer l'absence d'invasion des clones « pré-L » isolés à 5000 et 6000 générations par les clones S serait que ces clones « pré-L » ne soient pas encore capables de sécréter dans le surnageant la(les) substance(s) utilisée(s) par les S. Pour tester cette hypothèse, des filtrats de culture des clones 5000.5, 6000.2 et 6,5L4 ont été préparés et utilisés pour tester la croissance des clones 6K3 et 6,5S1 (Figure 36). A nouveau, les expériences ont été menées sur deux clones isolés indépendamment à 5000 et 6000 générations au sein de la ligne de descendance de la lignée L, et les résultats obtenus ont été similaires. Seuls les résultats obtenus pour 5000.5 et 6000.2 sont présentés ici. Les résultats ont révélé qu'à l'instar de l'ancêtre, les clones pré-L isolés à 5000 générations ne permettent pas la croissance des clones de type S et ne sécrètent donc aucun composé utilisable (Figure 36a). D'autre part, les résultats suggèrent que les relations syntrophiques entre L et S pourraient se mettre en place progressivement (Figure 36). En effet, alors qu'aucune croissance des clones 6K3 et 6,5S1 n'a été observée dans le surnageant de culture de 5000.5 (Figure 36a), une croissance faible a été détectée dans le filtrat de 6000.2 (Figure 36b), et une croissance complète dans le filtrat de 6,5L4 (Figure 36c).

Tous ces résultats impliquent, comme pour la lignée S, une histoire évolutive spécifique à la lignée L avec une succession de mutations conduisant à un génotype L. Cette histoire évolutive va ainsi amener à la divergence des lignées S et L via des relations syntrophiques et dépendantes de la fréquence. Nous avons à disposition les séquences de génomes du clone 5000.5, et le génome des clones 6000.2 et 6K3 est en cours de

réalisation. Les analyses de ces génomes permettront d'identifier des mutations candidates responsables de l'émergence de la lignée L et de ses caractéristiques propres. Les mêmes études que celles entreprises et réalisées pour les trois mutations *spoT*, *arcA* et *gntR* impliquées dans l'émergence de la lignée S pourront alors être effectuées. Nous pourrions alors tester un des points les plus intéressants, qui consisterait à comprendre si l'établissement successif des différentes mutations, au sein de chacune des lignées S et L, est réalisé de façon totalement indépendante ou si la succession des événements mutationnels d'une des lignées dépend de la succession des mutations de l'autre lignée. Il est d'ailleurs intéressant de noter une succession dans l'établissement des phénotypes caractéristiques de chacune des lignées S et L. Ainsi, ces différents résultats révèlent que la ligne de descendance menant à la lignée S se caractérise tout d'abord par l'apparition de la capacité à se développer sur le surnageant de cultures de clones amenant à la lignée L (Figure 36, clone 6K3), puis ensuite par la capacité à envahir les clones L grâce à des interactions négatives dépendant de la fréquence (Figures 34 et 35). D'autre part, les clones se situant sur la ligne de descendance de la lignée L acquièrent rapidement la capacité d'envahir les clones S, mais présentent de façon plus progressive la capacité de sécréter des substances utilisables par les S dans le milieu de culture.

Enfin, les interactions dépendantes de la fréquence mises en évidence contre l'ancêtre lui-même (Figure 33) constituent probablement un phénomène indépendant de celui mis en jeu dans l'émergence des lignées S et L et, bien que non expliquées à ce jour, représentent également un phénomène intéressant à étudier.

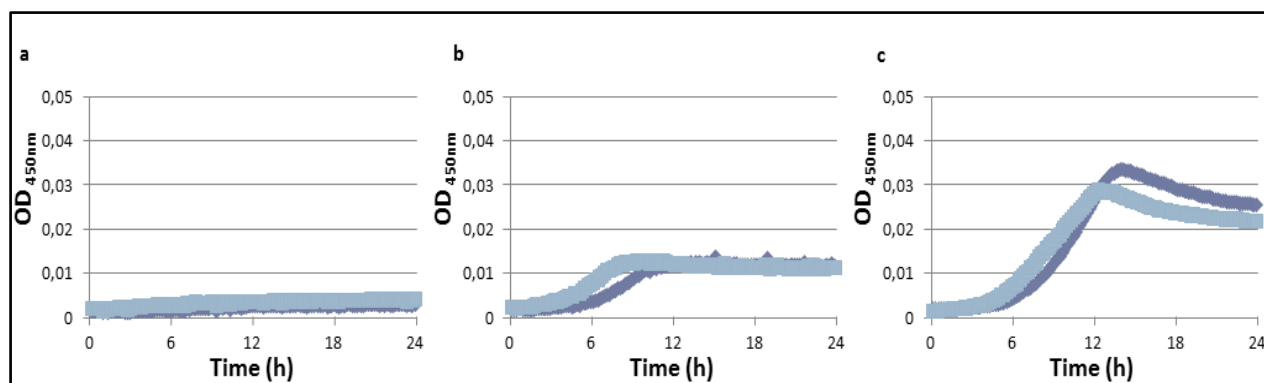


Figure 36 : Croissance des clones 6K3 et 6,5S1 dans le surnageant de culture des clones 5000.5 (a), 6000.2 (b) et 6,5L4 (c). Les clones 6K3 (bleu foncé) et 6,5S1 (bleu clair) ont été inoculés dans des filtrats de cultures issues des clones 5000.5 (a), 6000.2 (b) ou 6,5L4 (c) et leur croissance a été suivie. Les détails des expériences de croissance sur filtrat sont identiques à celles données dans l'article joint (Pluain et al).

RESULTATS PARTIE III

ANALYSES BIOCHIMIQUES ET

MOLECULAIRES DE LA

MUTATION DU GENE *arcA*

I. Introduction

Le système à deux composants ArcA/ArcB régule la transcription de plus d'une centaine de gènes chez *E. coli* en fonction des conditions d'oxygénation du milieu (voir Introduction Partie I section III). Il est composé de la protéine senseur ArcB et du régulateur de réponse ArcA (Malpica et al 2006). ArcA est une petite protéine de 29 kDa et 238 acides aminés, composée d'un domaine C-terminal contenant un motif hélice-tour-hélice de liaison à l'ADN (acides aminés 135 à 238) et d'un domaine N-terminal contenant le site de phosphorylation par ArcB (Toro-Roman et al 2005). Dans ce domaine N-terminal, le résidu Asp54 est le site de phosphorylation de ArcA, et la phosphorylation de ce résidu est nécessaire à la dimérisation de la protéine, qui elle-même est essentielle à la fixation de la protéine sur l'ADN (Toro-Roman et al 2005). Le domaine N-terminal de la protéine ArcA, des résidus 1 à 123, a été cristallisé en présence et en absence d'un analogue de groupement phosphoryle, BeF_3^- (Toro-Roman et al 2005). En plus du résidu Asp54 qui est phosphorylé, les résidus Thr81, Gly82 et Arg83 sont importants pour la stabilisation du groupement phosphoryle. Pour la dimérisation de ArcA, des interactions entre les résidus Ile90, Leu93, Ile112 d'un monomère et les résidus Lys89, Glu109 et Arg113 de l'autre sont notamment importantes, même si d'autres résidus sont également impliqués.

Une mutation dans le gène *arcA* a été mise en évidence au cours d'une expérience d'évolution en laboratoire où *E. coli* a été propagé en présence de deux sources de carbone, glucose et acétate (Le Gac et al 2010, voir Introduction Partie II section II.3). Cette mutation du gène *arcA* entraîne la modification du résidu Thr81 en Ala, ce résidu

étant impliqué dans la phosphorylation de ArcA. Cette mutation participe à un processus de diversification adaptative qui se produit dans ces conditions, bien que son impact moléculaire et physiologique soit inconnu.

Nous avons démontré au cours de ce travail (voir Résultats Partie II) qu'une mutation du gène *arcA* était impliquée directement dans l'événement de diversification adaptative qui s'est produit dans l'une des douze populations, Ara-2, initiées par Richard Lenski (Rozen et al 2000, 2005, 2009, Le Gac et al 2012). De plus, non seulement le gène *arcA* est important pour l'établissement de ce polymorphisme, mais également la nature précise de la mutation (voir Résultats Partie II). En effet, d'autres allèles du gène *arcA*, se produisant dans d'autres populations de l'expérience d'évolution au long terme, ne confèrent pas les caractéristiques propres à la lignée S. La mutation du gène *arcA* qui s'est produite dans la population Ara-2 entraîne le changement L93F au niveau de la protéine ArcA. Ce résidu acide aminé est connu pour être important pour la dimérisation de la protéine (Toro-Roman et al 2005). Afin d'étudier l'effet biochimique et moléculaire de cette mutation, nous avons purifié les protéines ArcA de la souche ancestrale et d'un clone S isolé de la population Ara-2. Ces deux protéines ont été analysées pour leur capacité à se lier à l'ADN par des expériences de retard sur gel avec des fragments d'ADN correspondant à deux gènes connus pour être régulés par ArcA, *rpoS* (Mika et Hengge 2005) et *fadE* (Cho et al 2006).

II. Matériel et Méthodes

1. Milieux de cultures, souches et plasmides

Les souches BW25113 *slyD::kan* (Baba et al 2006) et JM109 (Yanisch-Perron et al 1985) ont été utilisées pour surexprimer les protéines ancestrale (6His-ArcA-anc) et évoluée (6His-ArcA-S), respectivement. Les gènes codant les deux protéines ArcA ont été clonés sur le plasmide pQE80 (Qiagen[®]) qui permet l'introduction d'une étiquette de six résidus histidine (6His) à l'extrémité N-terminale des protéines codées par les gènes clonés. Le plasmide pQE80 a été utilisé pour les surexpressions des protéines ArcA. Après clonage dans pQE80, les gènes se retrouvent sous la dépendance du promoteur T5 auquel a été ajoutée la séquence de l'opérateur reconnu par le répresseur LacI. La transcription des gènes clonés peut donc être induite par l'ajout d'IPTG.

Les cellules bactériennes ont été cultivées en milieu riche Luria Broth (Sambrook et al 1989) à 37°C avec agitation. L'ampicilline a été utilisée à une concentration finale de 100 µg/mL et l'IPTG à 1mM, selon les recommandations d'usage (Qiagen[®]). La croissance a été suivie en mesurant la densité optique à 600 nm (OD₆₀₀).

2. Clonage des gènes *arcA* et surexpression des protéines ArcA

Pour obtenir les plasmides de surexpression pQE80-ArcA-anc et pQE80-ArcA-S, la région codante entière de chacun des gènes *arcA* de l'ancêtre et du clone évolué 6,5KS1 a été amplifiée par PCR en utilisant les amorces 5'-TATGGATCCCAGACCCCGCACATT-3' et 5'-GTCAAGCTTTTAATCTTCCAGATC-3',

qui ont été définies en introduisant un site de restriction pour les enzymes *Bam*HI et *Hind*III, respectivement. Chacun des deux fragments obtenus a été digéré par *Bam*HI et *Hind*III, puis cloné dans le vecteur pQE80 (Qiagen®) préalablement linéarisé par ces mêmes enzymes. Les inserts des plasmides recombinants obtenus ont été vérifiés par séquençage pour s'assurer de l'absence de mutations introduites au cours des étapes d'amplification PCR. Les plasmides recombinants pQE80-ArcA-anc et pQE80-ArcA-S ont été introduits par électro-transformation dans les bactéries *E. coli* BW25113 *slyD::kan* et JM109, respectivement. Les conditions de cultures des bactéries portant les plasmides de surexpression des protéines 6His-ArcA-anc et 6His-ArcA-S ont été celles décrites dans le protocole 8 de Qiagen (Qiagen®; <http://www.qiagen.com/literature/handbooks/literature.aspx?id=1000053>) et selon Mika et Hengge (2005). Brièvement, deux cultures indépendantes d'un litre chacune de LB contenant 25µg/mL de kanamycine ont été inoculées au 1/50^{ème} avec la souche portant le plasmide adéquat, puis incubées à 37°C avec agitation. A une DO_{600nm} de 0,6, l'expression de la protéine d'intérêt a été induite par l'ajout d'IPTG (concentration finale 1mM). Des prélèvements ont été effectués juste avant l'ajout d'IPTG, puis toutes les 2 ou 3 heures pour suivre l'expression de la protéine. Après 6 h d'incubation, les cultures ont été centrifugées, et le culot bactérien congelé à -20°C jusqu'à purification des protéines.

3. Purification des protéines ArcA

La purification des protéines 6His-ArcA-anc et 6His-ArcA-S a été réalisée selon les protocoles 9 et 12 développés par Qiagen (Qiagen®; <http://www.qiagen.com/literature/handbooks/literature.aspx?id=1000053>) pour la purification de protéines cytoplasmiques en conditions natives. Les tampons et solutions ont été préparés et utilisés selon ces protocoles (Qiagen®), à l'exception de la solution de phosphate de sodium qui a été remplacée par du tampon Tris 50mM pH8 comme décrit précédemment (Mika et Hengge 2005). Brièvement, après centrifugation des cultures et élimination des surnageants, les cellules ont été lysées par sonication dans un tampon contenant 20mM de β -mercaptoéthanol. Les protéines ont été purifiées par passage des extraits cellulaires sur colonnes de nickel. Les étiquettes histidine présentes dans la partie N-terminale des protéines ArcA permettent leur fixation sur la colonne de nickel, l'élution étant réalisée par une solution d'imidazole. L'imidazole du tampon d'élution a ensuite été éliminé par dialyse. Les protéines éluées ont ensuite été séparées par une étape de filtration sur gel, par passage sur une colonne Superdex-200 (GE Healthcare), qui permet une séparation des protéines en fonction de leur poids moléculaire. Les protéines ont enfin été concentrées en utilisant des filtres à centrifuger Centricon 10 U (Amicon) ou Vivaspin (Vivaspin), et congelées dans le tampon de stockage (Tris 10mM pH 8, EDTA 0,1mM, $MgCl_2$ 10mM, KCl 200mM, glycérol 50%). Les quantités de protéines obtenues ont été estimées sur la base d'un coefficient d'extinction 0,1% de 0,506 à 280nm (Nanodrop, Thermo Scientific).

Les protéines purifiées ont été analysées par spectrométrie de masse et séquençage N-terminal pour confirmer qu'il s'agissait bien des protéines ArcA. Pour la spectrométrie de masse, les protéines ont été lavées sur un macroTrap (C4, Michrom Bioresources) et une colonne RPHPLC (Jupiter-C4, 5u, 300Å 50x1.0mm, Phenomenex). Les échantillons ont été analysés avec un spectromètre de masse 6210 LC-TOF en interface avec un système de pompe LC (Agilent Technologies). Pour le séquençage N-terminal, la séquence en acides aminés a été déterminée sur la base de la réaction de dégradation d'Edman en utilisant un séquenceur en phase gazeuse (Applied Biosystems Model 492). Les procédures et réactifs ont été utilisés selon les protocoles recommandés par les fournisseurs.

4. Mesure des interactions ADN-protéine

Préparation des sondes : Les sondes ont été obtenues par PCR en utilisant les amorces ODS1063 : 5' cctcagaagcgccaaaggtt 3' et ODS1064 : 5' acgttggtcagacctgcag 3' pour la région promotrice du gène *rpoS*; et FadEF : 5' caggacttttgacctgaag 3' et FadER : 5' gacaaccgtagcgagaatac 3' pour la région promotrice du gène *fadE*, et l'ADN génomique de la souche *E. coli* REL606, qui est l'ancêtre de l'expérience d'évolution au long terme. Les sondes *rpoS* et *fadE* ont été marquées à la Digoxygénine (DIG, Roche Applied Bioscience), puis purifiées sur colonne (Macherey-Nagel). Le marquage par la Digoxygénine permettra des réactions d'immuno-détection, en utilisant des anticorps anti-Digoxygénine (voir plus bas).

Phosphorylation de ArcA : La réaction de phosphorylation des protéines 6His-ArcA-anc et 6His-ArcA-S a été effectuée dans le tampon EMSA (Tris-Cl 0,2M pH 7,4 ; KCl 0,2M ; MgCl₂ 0,02M et glycérol 4%), dans lequel ont été ajoutés 50mM d'acétyl phosphate, 20mM de DTT, 1 µL d'inhibiteur de protéases (cOmplete ULTRA tablet mini, Roche Applied Bioscience) et 10 µg de l'une ou l'autre des deux protéines. Les mélanges ont été incubés à 30°C pendant au moins 30 minutes.

Gel retard : Environ 140 ng de sonde (*rpoS* ou *fadE*), soit environ 0,05 µM, ont été incubés avec 0,9 µg de protéines (phosphorylées ou non) , soit environ 1,5 µM, en présence de tampon EMSA 1X et de DTT 2mM dans un volume final de 20µL, pendant 30 min à température ambiante. Les interactions protéines-ADN ont été visualisées après électrophorèse, dans des conditions natives, sur gel d'acrylamide 4% avec du tampon TBE (Tris Borate EDTA) 0,5X. La totalité de chacune des réactions protéines-ADN a été additionnée de 2µL de tampon de charge (tampon EMSA 1X, 4% glycérol, 2mM DTT, 0,2% bleu de bromophénol), puis déposée sur gel. Une étape de pré-migration de 45 min à 4°C et à 100V a été réalisée, suivie d'une migration à 4°C et à 300V. Un contrôle a été effectué avec de l'ADN froid (non marqué à la Digoxygénine), utilisé comme compétiteur spécifique.

Transfert sur membrane et immuno-détection : Les molécules présentes sur le gel ont été transférées sur une membrane Nylon N (GE Healthcare) en conditions humides (tampon TBE 0,5X ; 300mA pendant 1h15 à 4°C). Les acides nucléiques ont été fixés sur la membrane par exposition aux UV (Stratalinker, Stratagene). Les membranes ont ensuite subi des réactions d'immuno-détection en utilisant un anticorps anti-Digoxygénine couplé à la phosphatase alcaline (Roche Applied Bioscience). Les étapes

de saturation des membranes, d'hybridation, de lavages, et de révélation du marquage ont été effectuées selon le protocole du fournisseur (Roche Applied Bioscience). Après 5 min d'incubation dans le réactif CSPD, une exposition de 15 min de la membrane avec un film d'autoradiographie (Roche Applied Bioscience) a été effectuée.

III. Résultats

1. Surexpression de ArcA

La souche JM109 a tout d'abord été transformée par chacun des deux plasmides pQE80-ArcA-anc et pQE80-ArcA-S pour surexprimer les protéines 6His-ArcA-anc et 6His-ArcA-S. La surexpression de la protéine 6His-ArcA-S n'a pas influencé la croissance des cellules bactériennes et les quantités de protéine surexprimée se sont révélées importantes (Figure 37b). Nous avons donc poursuivi la purification de cette protéine (voir Matériel et Méthodes et section suivante). Le poids moléculaire de la protéine est conforme à celui attendu (28,5 kDa).

Des difficultés plus importantes ont été rencontrées dans le cas de la protéine 6His-ArcA-anc. En effet, sa surexpression a entraîné un taux de croissance faible de JM109, une DO maximum faible, et des quantités de protéine recombinante également faibles (Figure 37a). Ces caractéristiques sont souvent liées à une toxicité de la protéine surexprimée. Nous avons donc modifié les conditions de cultures pour la surexpression, ces modifications étant recommandées (Qiagen®) en cas de toxicité de la protéine. En particulier, la température d'incubation des cultures a été diminuée, le moment d'induction à l'IPTG retardé, et les concentrations en IPTG modifiées. Ces modifications

n'ont cependant conduit à aucune amélioration des quantités de protéine 6His-ArcA-anc produite (résultats non montrés). Nous avons donc décidé de tout de même poursuivre la purification de cette protéine, en utilisant les mêmes conditions que pour la protéine 6His-ArcA-S qui n'a pas posé de problème de toxicité. Le poids moléculaire de la protéine est conforme à celui attendu (28,5 kDa).

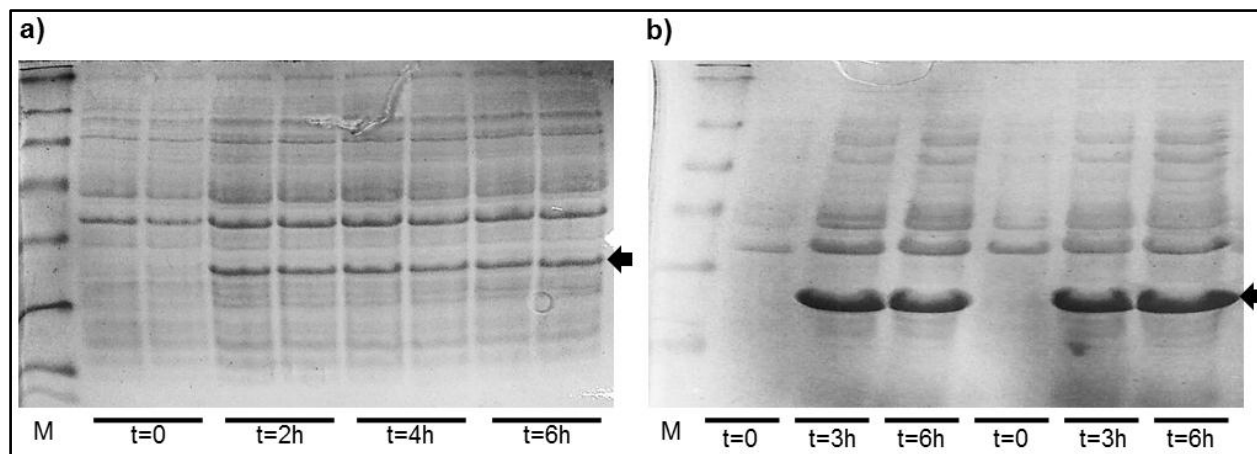


Figure 37 : Surexpression des protéines 6His-ArcA-anc (a) et 6His-ArcA-S (b). Des cultures d'*E. coli* JM109 contenant chacun des deux plasmides pQE80-ArcA-anc et pQE80-ArcA-S ont été réalisées et à une DO_{600nm} de 0,6, de l'IPTG a été rajouté pour induire l'expression des protéines 6His-ArcA-anc et 6His-ArcA-S ($t=0$). Des prélèvements ont été effectués juste après l'ajout d'IPTG ($t=0$) et à différents temps par la suite. Des quantités égales de protéines totales ont été déposées sur gel pour chaque temps de prélèvement. Chaque expérience de surexpression a été réalisée en duplicata. Les protéines ArcA sont montrées par des flèches. M est un mélange de protéines dont le poids moléculaire est connu et qui constitue donc un marqueur de taille.

2. Purification des protéines ArcA

La purification des protéines 6His-ArcA-anc et 6His-ArcA-S surexprimées a été effectuée par passage sur colonnes de nickel, suivi d'une filtration sur gel (voir Matériel et Méthodes). Comme attendu après la surexpression, des quantités importantes de protéine 6His-ArcA-S pure ont pu être obtenues (Figure 38b et d). La pureté de la protéine a été vérifiée par spectrométrie de masse et séquençage N-terminal.

Malgré les problèmes rencontrés au cours de la surexpression de la protéine 6His-ArcA-anc, des quantités faibles, mais suffisantes, de la protéine ont pu être obtenues après passage sur colonne de nickel et gel filtration (Figure 38a et c). Cependant, le séquençage N-terminal de la protéine purifiée 6His-ArcA-anc a révélé que celle-ci était contaminée par une protéine identifiée comme étant SlyD. Cette protéine d'*E. coli* possède naturellement 14 résidus His dans sa partie C-terminale, ces résidus constituant ainsi une étiquette histidine naturelle, qui lui a permis de s'accrocher sur la colonne de nickel et d'être ainsi co-éluee en même temps que la protéine recombinante 6His-ArcA-anc. De plus, la protéine SlyD est de taille similaire à 6His-ArcA-anc et ne peut donc pas être séparée lors de l'étape de filtration sur gel (Robichon et al 2011). Pour éviter cette complication et obtenir une protéine pure, la surexpression de la protéine 6His-ArcA-anc a été ré-initiée, mais cette fois en introduisant le plasmide de surexpression dans un mutant d'*E. coli* délété du gène *slyD*. Cette fois, la protéine 6His-ArcA-anc a pu être obtenue pure, mais toujours en quantités faibles (résultats non montrés). La contamination par la protéine SlyD n'a pas été détectée lors de la surexpression et de la purification de la protéine 6His-ArcA-S, de par la quantité importante de protéine 6His-ArcA-S produite.

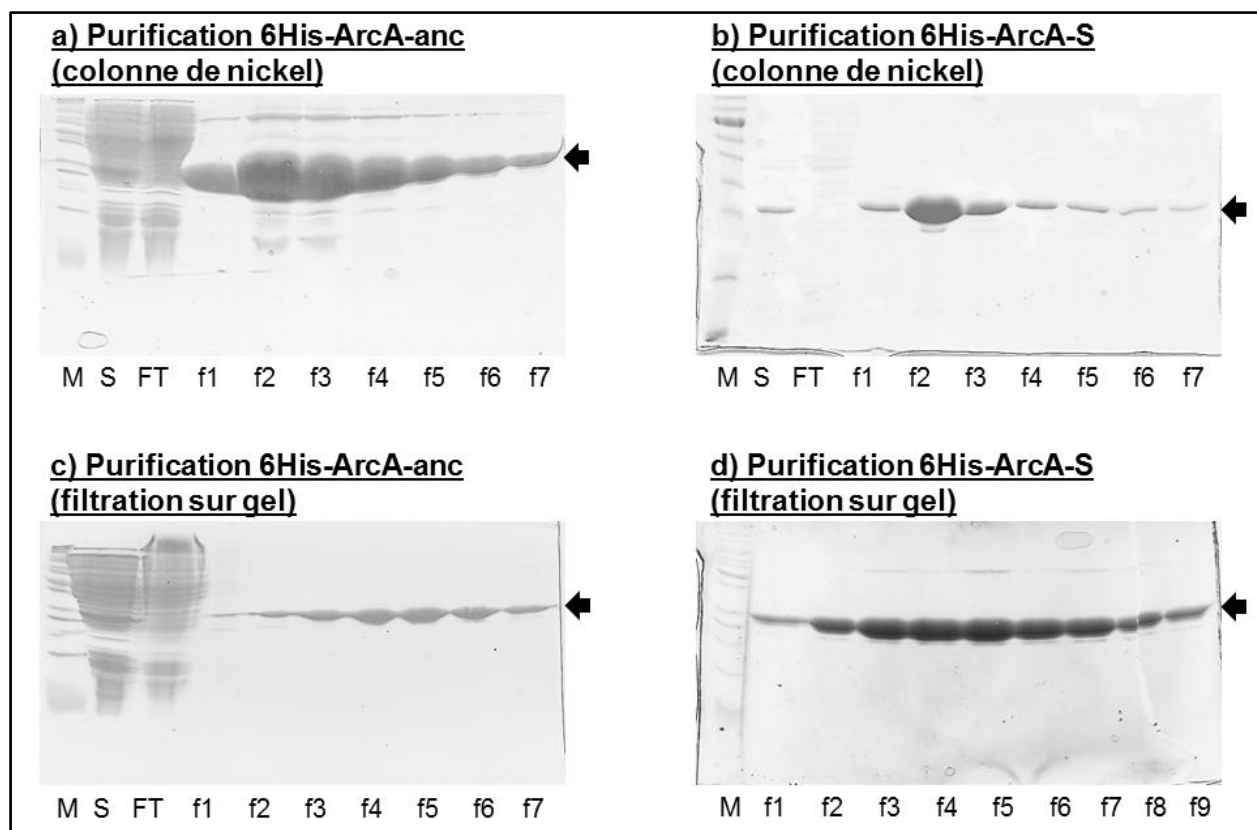


Figure 38 : Purification des protéines 6His-ArcA-anc (a et c) et 6His-ArcA-S surexprimées (b et d). Après purification sur colonne de nickel (a et b), puis par filtration sur gel (c et d), un volume équivalent (10 μ L) des différentes fractions récupérées a été déposé sur gel d'acrylamide pour vérifier le profil protéique. M : marqueurs de poids moléculaire ; S : surnageant de centrifugation après lyse des cellules ; FT (Flow-Through) : protéines non retenues sur la colonne de nickel ; f1 à f9 : numéro des fractions récupérées et déposées sur gel. Les flèches indiquent la position de la protéine ArcA.

3. Interactions protéines-ADN

La protéine ArcA est un régulateur de la transcription de plus d'une centaine de gènes chez *E. coli* (Salmon et al 2005). La protéine ArcA se fixe, après phosphorylation par ArcB et sous forme de dimères, sur un site spécifique présent dans la région promotrice des gènes régulés directement, dont *rpoS* (Mika et Hengge 2005) et *fadE* (Cho et al 2006). Afin de déterminer si le changement d'acide aminé provoqué par la mutation, identifiée chez les clones de la lignée S et impliquée dans leur émergence, affectait la

fonction de fixation à l'ADN de la protéine ArcA, nous avons procédé à des mesures d'interactions protéines-ADN par des expériences de retard sur gel. Les protéines 6His-ArcA-anc et 6His-ArcA-S ont tout d'abord été phosphorylées *in vitro*, car la phosphorylation est essentielle à l'activité de fixation de ArcA à l'ADN. Les protéines phosphorylées ont alors été mises en présence de la région promotrice de chacun des gènes *rpoS* et *fadE* qui présentent des sites de fixation de ArcA. Des expériences contrôles, sans protéines ou avec les protéines non phosphorylées, ont été incluses, ainsi qu'une expérience avec un compétiteur spécifique mais non marqué (voir Matériel et Méthodes).

Les résultats montrent que la protéine 6His-ArcA-anc sous sa forme phosphorylée induit un retard de migration de l'ADN pour chacune des deux sondes (Figure 39), ce qui confirme sa fixation sur la région promotrice de *rpoS* et *fadE*. Comme attendu, le contrôle avec le compétiteur spécifique annule le retard (résultats non montrés). Contrairement à la protéine ancestrale, la protéine 6His-ArcA-S n'induit pas ce retard de migration de l'ADN (Figure 39), que ce soit sous sa forme phosphorylée ou non. La protéine évoluée n'est donc plus capable, ou trop faiblement, de se fixer sur la région promotrice des gènes *rpoS* et *fadE*. Cette différence de fixation entre les deux protéines ancestrale et évoluée est beaucoup plus accentuée dans le cas de la région promotrice du gène *rpoS*.

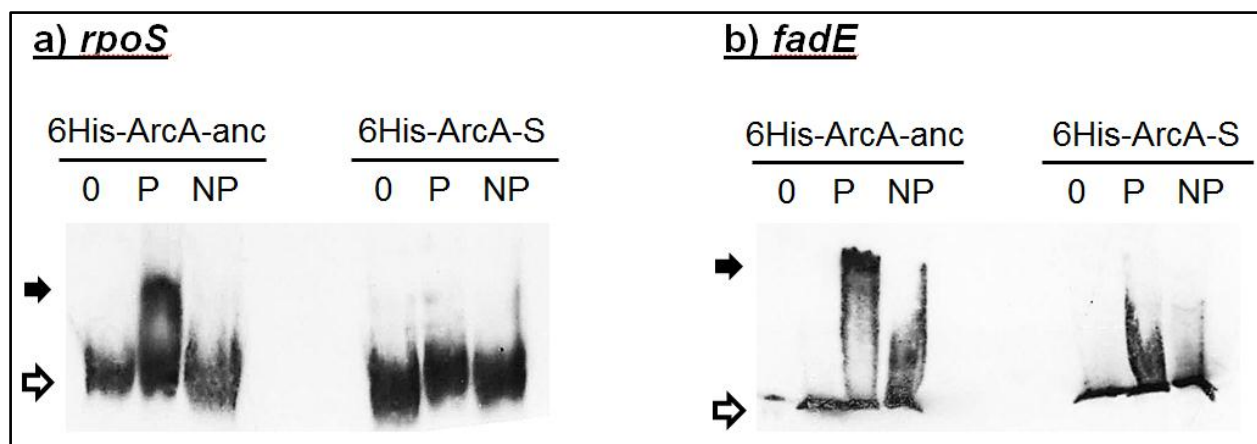


Figure 39 : Fixation des protéines 6His-ArcA-anc et 6His-ArcA-S sur les régions promotrices de *rpoS* (a) ou *fadE* (b). Les sondes *rpoS* (a) et *fadE* (b), marquées à la Digoxygénine, ont été incubées en absence de protéine (0), et en présence de la protéine préalablement phosphorylée (P) ou non (NP). La protéine utilisée est indiquée. Les expériences ont toutes été réalisées dans les mêmes conditions. Un contrôle en présence d'ADN non marqué (compétiteur spécifique) a été réalisé à chaque fois, et annule le retard observé (résultats non montrés). Les flèches creuses indiquent les fragments non retardés, et les flèches pleines les fragments retardés par la fixation des protéines.

IV. Conclusions

La mutation L93F détectée chez les clones de la lignée S et impliquée dans son émergence, engendre donc une modification des capacités de fixation de la protéine ArcA sur au moins deux des promoteurs qu'elle régule. Ces expériences doivent maintenant être complétées tout d'abord afin de quantifier ces différences. En effet, les expériences de retard sur gel seront réalisées en présence de quantités croissantes des protéines 6His-ArcA-anc et 6His-ArcA-S, afin d'estimer les constantes d'affinité de chacune pour l'ADN. D'autre part, deux hypothèses peuvent être envisagées pour expliquer la perte, ou la forte diminution, de la capacité de la protéine évoluée à lier l'ADN. En effet, la protéine 6His-ArcA-S pourrait être affectée, soit dans sa capacité à être phosphorylée, soit dans sa faculté de dimérisation. Ces deux hypothèses pourront être testées directement. Nous avons ainsi également purifié au cours de ce travail la

protéine ArcB (résultats non montrés). Elle pourra ainsi être utilisée pour examiner la phosphorylation des protéines ArcA ancestrale et évoluée, et le niveau de phosphorylation pourra alors être quantifié. Les propriétés de dimérisation pourront quant à elles être testées par des expériences de dichroïsme circulaire, en collaboration avec l'Institut de Biologie Structurale de Grenoble.

La perte de la capacité de la protéine 6His-ArcA-S de lier l'ADN pourrait influencer ses propriétés de régulation de l'expression des gènes. Cet effet potentiel peut être testé de différentes manières. Tout d'abord, la transcription de gènes cibles, connus pour être régulés par ArcA (Salmon et al 2005), pourra être analysée par RT-PCR quantitative à partir d'ARNs extraits de différentes souches : l'ancêtre, un clone S isolé à 6500 générations par exemple, et l'ancêtre dans lequel nous avons introduit la mutation du gène *arcA*. D'autre part, l'impact de la mutation du gène *arcA* pourra être analysé de façon globale sur la transcription de l'ensemble des gènes par des expériences de transcriptome. Les profils globaux de transcription de l'ancêtre, de clones S et L, ainsi que de souches isogéniques que nous avons reconstruites (voir Résultats Partie II), pourront ainsi être comparés. Ces données devraient permettre non seulement de comprendre l'impact de la mutation du gène *arcA* sur la transcription de l'ensemble de ses gènes cibles, mais aussi de mettre en évidence d'éventuels effets épistatiques c'est-à-dire des effets différentiels de cette mutation en fonction du contexte génétique dans lequel elle se trouve. Ces expériences pourraient notamment permettre d'identifier les mécanismes moléculaires qui relient l'altération du gène *arcA* à l'événement de diversification adaptative mis en évidence dans la population Ara-2.

Seul l'allèle mutant substitué dans la population Ara-2 engendre l'opportunité conduisant à l'émergence de la lignée S (voir Résultats Partie II). En effet, les allèles évolués du gène *arcA*, mis en évidence dans d'autres des douze populations d'évolution, ne conduisent pas aux phénotypes caractéristiques des clones S. Les mécanismes à la base de cette spécificité d'allèles sont encore inconnus. Pour essayer de les comprendre, nous allons surexprimer et purifier, par des techniques similaires à celles utilisées dans ce chapitre, les protéines ArcA évoluées issues des autres populations d'évolution. Ces protéines pourront alors être utilisées dans des expériences de retard sur gel pour analyser si elles sont également altérées pour leur capacité de fixation à l'ADN. Les profils globaux de transcription, associés à ces autres allèles de *arcA*, pourraient également permettre de savoir si les mêmes gènes cibles sont affectés par les différents allèles évolués du gène *arcA*. L'objectif ultime est ainsi de comprendre cet événement de diversification adaptative du niveau écologique au niveau moléculaire.

DISCUSSION

GENERALE

Ces travaux ont établi l'originalité et l'intérêt des stratégies d'évolution expérimentale pour l'étude approfondie des événements de diversification. En effet, ces événements complexes, qui mènent à la co-existence de types bactériens différenciés, sont difficiles à étudier, notamment au niveau génomique, dans les environnements naturels, car il s'agit fréquemment de processus uniques s'étant produits dans un passé lointain. En revanche, lorsque de tels événements se produisent au cours d'expériences d'évolution en laboratoire, dont le principe même est de pouvoir disposer d'archives fossiles vivantes incluant l'ancêtre et tous les intermédiaires évolutifs, ils peuvent être disséqués de façon rigoureuse grâce aux données physiologiques et génétiques comparatives issues de ces « strates fossiles vivantes ».

Au cours d'expériences d'évolution en laboratoire, des événements de diversification se produisent notamment dans des environnements structurés, au sein desquels différentes niches écologiques sont disponibles et accessibles à différents types bactériens issus d'un ancêtre commun (voir Introduction Partie II). Dans ces conditions, l'émergence de cette diversification se fait de façon répétée. L'originalité et l'intérêt de l'événement de diversification étudié au cours de cette thèse résultent d'une part du fait qu'il s'agit d'un processus unique au sein de populations indépendantes, non attendu *a priori* dans ces conditions expérimentales non structurées qui visaient justement à essayer de réduire la complexité, et d'autre part du fait que le polymorphisme qui en résulte est le plus long et le plus stable jamais observé à ce jour dans ces expériences d'évolution en laboratoire. L'étude, tant aux niveaux physiologique, écologique que génétique et génomique de l'émergence et du maintien de ce polymorphisme, peut ainsi

s'avérer précieuse pour la compréhension des mécanismes généraux de diversification bactérienne.

Les résultats obtenus au cours de ce travail nous ont notamment permis de déterminer les bases génétiques de l'émergence d'un polymorphisme stable, qui ont révélé l'importance de trois processus évolutifs majeurs dans cet événement de diversification, que nous allons maintenant discuter.

Premièrement, la restructuration des réseaux de régulation géniques et métaboliques semble prépondérante dans l'émergence du polymorphisme, puisque les trois mutations impliquées affectent toutes trois des gènes codant des régulateurs de la transcription - *spoT*, *arcA* et *gntR* – dont les gènes cibles sont notamment impliqués dans le métabolisme central de la bactérie.

Deuxièmement, la spécificité de l'allèle des gènes *spoT* et *arcA* est un des paramètres clé expliquant l'unicité de l'événement de diversification, qui s'est ainsi produit uniquement dans la population Ara-2, sans aucun parallélisme phénotypique détecté dans les onze autres.

Troisièmement, l'émergence du polymorphisme est dépendante d'une succession de trois événements mutationnels, au moins dans la lignée S.

I. Plasticité des réseaux de régulation et des réseaux métaboliques

Les bactéries ont colonisé pratiquement toutes les niches environnementales de la biosphère. Elles peuvent ainsi utiliser une très large variété de sources de carbone et/ou

d'énergie, vivre dans des conditions variées de concentration en oxygène et interagir avec de multiples autres organismes vivants. Ainsi et de façon peu surprenante, les gènes codant des enzymes du métabolisme font partie des cibles majeures de la sélection naturelle au cours d'expériences d'évolution en laboratoire (Dykhuizen et Dean 1990, Cooper et al 2001, Herring et al 2006, Woods et al 2006, Lee et Palsson 2010, Khan et al 2011, Hindré et al 2012). D'autre part, des changements globaux d'expression de gènes impliqués dans le métabolisme central ont également été détectés lors de telles expériences, suggérant des modifications de la régulation des voies métaboliques (Ferea et al 1999, Gresham et al 2008, Le Gac et al 2008). Les études génétiques appliquées aux expériences d'évolution en laboratoire, ainsi que la possibilité d'accéder à des séquences entières de génomes, ont ainsi révélé des mutations affectant des gènes codant des régulateurs globaux, dont certaines des cibles sont impliquées dans le métabolisme bactérien (Herring et al 2006, Le Gac et al 2008, Barrick et al 2009, Conrad et al 2010, Hindré et al 2012).

Quelles que soient les capacités métaboliques des bactéries, le cœur de leurs activités métaboliques est constitué par le métabolisme central, dont certains intermédiaires sont des précurseurs permettant à la fois les voies de biosynthèse conduisant *in fine* à la production de biomasse et la synthèse d'énergie sous forme d'ATP. Deux des trois mutations identifiées au cours de ce travail, nécessaires et suffisantes pour produire les phénotypes caractéristiques de la lignée S, affectent des gènes ayant un rôle majeur dans la régulation du métabolisme central, en l'occurrence *arcA* et *gntR*. La protéine ArcA régule directement les gènes codant des enzymes du cycle de Krebs et du shunt

du glyoxylate (Iuchi et al 1988, 1989), et GntR ceux codant les enzymes de la voie alternative à la glycolyse, la voie Entner-Doudoroff (Conway 1992). Ce sont donc les deux parties du métabolisme central, haute et basse, dont la régulation est modifiée dans la lignée S. Le cycle de Krebs constitue la partie basse du métabolisme central (Neidhart 1996), qui conduit à l'oxydation de l'acétyl-CoA provenant du pyruvate produit par la glycolyse. Le cycle de Krebs conduit alors à la production de précurseurs pour la synthèse de plusieurs acides aminés et des molécules à fort pouvoir réducteur, comme le NADH, dont les atomes d'hydrogène vont servir à la production d'ATP via la chaîne respiratoire. D'autre part, il existe dans le cycle de Krebs une voie de contournement appelée shunt du glyoxylate, qui est essentielle pour la croissance sur des sources de carbone alternatives telles que l'acétate (Kornberg 1966). La voie Entner-Doudoroff quant à elle, se situe dans la partie haute du métabolisme central et constitue une voie alternative à la glycolyse, conduisant à la dégradation du glucose en pyruvate (Conway 1992). Il s'agit d'une voie mineure, très peu utilisée par *E. coli* dans des conditions classiques de culture, mais qui semble activée dans le gros intestin où des sucres comme le gluconate, le glucuronate, ou le galacturonate, qui sont les substrats naturels de cette voie, sont présents en quantité abondante (Peekhaus et Conway 1998). Cette voie fait appel à deux enzymes, Edd et Eda, dont l'expression est réprimée par GntR. Au vu des fonctions pleiotropes complexes remplies par les protéines ArcA et GntR, il est difficile d'imaginer précisément l'effet physiologique des mutations conduisant aux phénotypes (valeur sélective, interactions dépendantes de la fréquence, capacités d'invasion) de la lignée S et donc à l'émergence du polymorphisme. De plus, nos résultats montrent que leurs effets dépendent du contexte génétique dans lequel elles

sont substituées (voir Résultats Partie II). Au niveau moléculaire, nos résultats montrent cependant que la mutation de *arcA* affecte la capacité de la protéine à se fixer sur l'ADN, au moins au niveau de deux gènes cibles (voir Résultats Partie III). Bien que l'effet de cette mutation soit inconnu sur l'expression globale des gènes, ces résultats suggèrent un relâchement de l'activité régulatrice de la protéine ArcA. Celle-ci étant un répresseur de la transcription des gènes codant les enzymes du cycle de Krebs et du shunt du glyoxylate, un effet potentiel pourrait être une augmentation de l'activité de ces derniers, même si cela doit être confirmé par la mesure des activités spécifiques des enzymes correspondantes. Cette augmentation de l'activité du cycle de Krebs et du shunt du glyoxylate devrait conduire à un déséquilibre du métabolisme central, notamment au niveau des voies haute et basse, ce qui devrait être défavorable aux cellules bactériennes (Waegeman et al 2011). Ces effets pourraient ainsi expliquer les tailles plus petites des colonies de la lignée S. Ceci pourrait être contre-balancé par d'autres phénotypes conférés suite à la mutation du gène *arcA*. En effet, la substitution de cette mutation est concomitante à la capacité des clones S de se développer sur un surnageant de culture des clones L. Ce phénotype pourrait alors conférer un avantage suffisant dans ces conditions de la population Ara-2 pour contre-balancer les effets néfastes potentiels de la mutation *arcA*. Ainsi, la restructuration métabolique associée à la modification du gène *arcA* pourrait être à la base de cette innovation phénotypique.

La mutation du gène *gntR* entraîne probablement l'inactivation (ou une baisse d'activité) de la protéine GntR. En effet, contrairement aux gènes *spoT* et *arcA*, aucune spécificité d'allèle n'a été détectée pour le gène *gntR*, la mutation substituée dans le même gène mais dans une autre population provoquant les mêmes phénotypes. Bien que cela doive

être confirmé par l'analyse de la protéine GntR évoluée et par des mesures d'activités des enzymes de la voie Entner-Doudoroff, l'effet de l'allèle évolué *gntR* pourrait donc être de déréguler cette voie alternative à la glycolyse. Il est ainsi intéressant de noter que la mutation du gène *gntR* est postérieure à celle de *arcA* dans la lignée S. Cette mutation pourrait ainsi, en modifiant la partie haute du métabolisme central, compenser les effets néfastes potentiels associés à la mutation du gène *arcA*, discutés ci-dessus. Cependant, les mécanismes à la base de ces effets sur la réorganisation du métabolisme central restent inconnus. Enfin, nos résultats montrent que l'introduction de la mutation *gntR* confère, dans le contexte génétique du clone évolué 6K3 qui porte les allèles évolués des gènes *spoT* et *arcA*, la dernière étape phénotypique conduisant aux interactions dépendantes de la fréquence (voir Résultats Partie II). La flexibilité du métabolisme central est donc à la base des capacités d'invasion réciproque et de l'émergence de ce polymorphisme.

La première mutation conduisant à la diversification entre les lignées S et L affecte le gène *spoT*, qui est central dans un des réseaux de régulation de l'expression des gènes les plus globaux, la réponse stringente (voir Introduction Partie I Section II.1). Une mutation de ce gène a donc des effets extrêmement pleiotropes car le gène *spoT*, en contrôlant le métabolisme de (p)ppGpp, est impliqué dans la reprogrammation de la machinerie transcriptionnelle dans la cellule. De plus, seule la mutation du gène *spoT* apparue dans la population Ara-2 confère les phénotypes caractéristiques de la lignée S, lorsqu'elle est combinée aux mutations des gènes *arcA* et *gntR*. Sur les trois mutations nécessaires aux innovations phénotypiques de la lignée S, celle dans *spoT*

est également la seule substituée lors de la période pré-divergence et elle est donc présente dans les deux lignées S et L. Pour toutes ces raisons, les effets de l'allèle évolué *spoT* sont difficiles à prédire et peuvent avoir des conséquences à différents niveaux. L'allèle évolué de la population Ara-2 pourrait entraîner une modification des quantités de (p)ppGpp dans la cellule, ce qui provoquerait des changements de régulation de certains gènes cibles. La réponse stringente étant impliquée dans les transitions entre états nutritionnels, et donc dans l'adaptation aux transitions entre phases exponentielle et stationnaire, la modification du gène *spoT* pourrait alors offrir une opportunité d'adaptation différentielle à l'une ou l'autre de ces phases du cycle cellulaire. Il a ainsi été démontré que les lignées L et S coexistaient grâce à des bénéfices différentiels, en phase exponentielle de croissance et en phase stationnaire, respectivement (Rozen et al 2009). Les mutations apparaissant par la suite, et spécifiques de chacune des deux lignées, pourraient alors renforcer ces différences dans ce contexte génétique affecté pour le gène *spoT*.

Ainsi, la réorganisation profonde des réseaux de régulation globale et du métabolisme central a fourni la flexibilité nécessaire à l'émergence d'une innovation phénotypique majeure, c'est-à-dire l'apparition de deux types cellulaires différenciés qui colonisent des niches écologiques différentes et co-existent lors du polymorphisme le plus stable mis en évidence dans des expériences d'évolution. Cela étend considérablement les connaissances et les propriétés connues pour ces réseaux cellulaires. En effet, ils sont donc non seulement impliqués dans la réponse immédiate des cellules bactériennes aux différentes conditions de stress rencontrées dans leur environnement, mais ils peuvent

être également remodelés à plus long terme pour assurer l'aptitude à la colonisation de nouvelles niches.

II. Spécificité allélique et unicité de l'événement de diversification

Les expériences d'évolution, au cours desquelles des populations indépendantes issues d'un ancêtre commun sont propagées dans des environnements identiques, se caractérisent par un très fort niveau de parallélisme phénotypique et génétique (Zambrano et al 1993, Wichman et al 1999, Zinser et Kolter 2000, Cooper et al 2001, Ferenci 2001, Cooper et al 2003, Seeto et al 2004, Crozat et al 2005, Pelosi et al 2006, Woods et al 2006, Cooper et al 2008, Philippe et al 2009, Crozat et al 2010, Hindré et al 2012). Ce parallélisme a également été détecté au niveau de l'émergence d'événements de diversification soit dans des environnements hétérogènes (Spencer et al 2007, McDonald et al 2009, voir Introduction Partie II), soit par construction de niche dans des environnements homogènes (Rosenzweig et al 1994, voir Introduction Partie II). Dans ces cas de figure, une seule mutation suffit à expliquer l'apparition de nouveaux types cellulaires (Treves et al 1998, Bantinaki et al 2007, Spencer et al 2007, McDonald et al 2009), ce qui explique probablement en partie le parallélisme observé dans les cas décrits de diversification.

En revanche, malgré ce niveau élevé de parallélisme génétique qui se caractérise par des mutations bénéfiques affectant souvent les mêmes gènes, les positions nucléotidiques modifiées et la nature des mutations sont différentes dans la grande majorité des cas au sein des populations indépendantes propagées lors d'une même

expérience d'évolution (Treves et al 1997, Woods et al 2006, Bantinaki et al 2007, Crozat et al 2010). Ceci génère ainsi une forte divergence allélique dont les effets ont été étudiés dans au moins deux cas, qui suggèrent que les différents allèles d'un même gène cible de la sélection naturelle tendent à converger vers un fort parallélisme moléculaire. Dans le cas des lignées de *P. fluorescens* propagées en environnement spatialement structuré (voir Introduction Partie II Section II.2), de nombreuses mutations différentes affectent des gènes communs des voies régulatrices Wsp, Aws et Mws, qui contrôlent la production de c-di-GMP et, en conséquence, de cellulose. Les différents allèles des mêmes gènes, bien que pouvant avoir des effets différents sur la valeur sélective (McDonald et al 2009), convergent toutes vers une modulation de la quantité de c-di-GMP qui entraîne une forte augmentation de la production de cellulose pour *in fine* aboutir à l'apparition du phénotype WS. Le second cas concerne l'expérience d'évolution au long terme chez *E. coli*, où un opéron, contenant les gènes *fis* et *dusB* et impliqué dans le contrôle du niveau de superhélicité de l'ADN, est la cible de la sélection naturelle de façon répétée, mais avec des variations alléliques entre les différentes populations. De façon similaire, alors que les différents allèles peuvent conférer des valeurs sélectives différentes, ils convergent tous vers une modulation comparable du niveau de superhélicité de l'ADN bactérien. Ainsi, les divergences alléliques conduisent à des effets moléculaires convergents, mais à des effets sur la valeur sélective qui peuvent être différents.

L'événement de diversification identifié dans la population Ara-2 se caractérise par son unicité. En effet, aucune des onze autres populations de cette expérience d'évolution ne

présente ce phénomène, même si une autre population présente une divergence avec apparition de lignées différentes. Dans ce second cas cependant, la diversification est réalisée sur la base de phénotypes totalement différents, ici l'utilisation ou non du citrate comme source de carbone (Blount et al 2008, 2012), et ce polymorphisme n'est que transitoire. L'unicité de l'événement de diversification conduisant aux lignées S et L pourrait être liée à deux caractéristiques génétiques majeures : il s'agit d'une part d'un phénomène complexe, nécessitant trois mutations conduisant à la lignée S et un nombre encore inconnu de mutations conduisant à la lignée L, ce qui rend la probabilité d'un tel événement faible. D'autre part, une très forte spécificité d'allèles au niveau des gènes *spoT* et *arcA* est associée à l'émergence de cet événement de diversification. En effet, seul l'allèle des gènes *spoT* et *arcA* substitué dans la population Ara-2 peut conduire aux phénotypes de la lignée S.

Les gènes *spoT*, *arcA* et *gntR*, dont la modification est à l'origine des innovations phénotypiques conduisant à la lignée S, sont les cibles de la sélection naturelle dans d'autres populations : 8 populations sur les 12 pour *spoT*, les 12 pour *arcAB*, et 2 pour *gntR*. Nos résultats ont montré que les allèles évolués des gènes *spoT* et *arcA* identifiés dans la population Ara-2 confèrent spécifiquement les interactions dépendantes de la fréquence qui caractérisent les lignées S et L. *A contrario*, l'allèle évolué *gntR* de la population Ara-2 peut être interchangé avec celui identifié dans une autre population sans interférer avec les phénotypes conférés. Si non seulement les gènes, mais également la nature des allèles mutés s'avèrent importants pour le développement de phénotypes donnés, cela restreint bien évidemment les opportunités de leur émergence. Il est cependant également possible de considérer ces résultats sous un angle différent.

En effet, ils montrent que des modifications alléliques variées, au sein de gènes régulateurs, génèrent des opportunités évolutives multiples et complexes. Alors que différents allèles de gènes régulateurs produisent des effets moléculaires semblables, les conséquences phénotypiques, incluant la valeur sélective, peuvent diverger. Comme nous l'avons vu ci-dessus, les réseaux de régulation confèrent une flexibilité immédiate aux bactéries en leur permettant de répondre aux différentes conditions de stress qu'elles peuvent rencontrer, mais également à plus long terme au cours de l'évolution. Nos résultats montrent un aspect supplémentaire de cette flexibilité, de par l'étendue des possibilités conférées par des allèles différents de gènes régulateurs. Cette multiplicité de phénotypes potentiellement conférés par des allèles différents révèle la pluralité des fonctions sous-jacentes contrôlées par les réseaux de régulation. Ainsi, des résultats préliminaires obtenus au sein de notre équipe suggèrent que plus de 150 allèles différents du gène *spoT* peuvent être sélectionnés pour leurs effets bénéfiques potentiels dans l'environnement de l'expérience d'évolution au long terme, c'est-à-dire un milieu minimum avec du glucose en quantité limitante comme seule source de carbone.

Devant les contraintes génétiques fortes qui pèsent sur l'émergence de l'événement de diversification, la question pourrait même se poser de savoir pourquoi un tel événement de diversification a finalement pu être détecté. De plus, ces contraintes fortes ont pu être déterminées uniquement pour la lignée S, mais les mêmes pèsent probablement sur l'émergence de la lignée L. Ainsi, nous pouvons par exemple nous interroger sur l'impact du statut mutateur de la population Ara-2 sur la probabilité d'émergence du

polymorphisme. Le phénotype mutateur apparaît très précocement dans cette population, vers 3000 générations, suite à une mutation dans le gène *mutL* impliqué dans la réparation des mésappariements de l'ADN. Les contraintes génétiques nécessaires à l'émergence du polymorphisme auraient-elles pu être surmontées si la population avait conservé un taux de mutation ancestral ? Cette question pourrait par exemple être posée en rejouant le film de l'évolution de la population Ara-2 à partir de différents intermédiaires évolutifs dans lesquels l'allèle ancestral du gène *mutL* aurait été rétabli.

III. Succession de mutations

La succession de changements au cours des processus évolutifs est à la base de la théorie de l'évolution (Darwin 1859). Les expériences d'évolution en laboratoire révèlent plusieurs exemples qui montrent que les trajectoires génétiques, ou chemins mutationnels, ne sont pas tous équivalents. Ainsi, l'ordre d'apparition des mutations est important au cours des processus adaptatifs. Au cours d'une étude récente, les cinq premières mutations bénéfiques substituées dans la population Ara-1 de l'expérience d'évolution au long terme ont été combinées dans tous les ordres possibles dans le contexte génétique de l'ancêtre (Khan et al 2011). Les différents chemins mutationnels reconstruits ne présentent pas une augmentation similaire de la valeur sélective et des interactions épistatiques entre certaines de ces cinq mutations ont été mises en évidence. De plus, une de ces cinq mutations a été substituée dans la population, non seulement parce qu'elle confère un effet positif direct sur la valeur sélective, mais aussi

suite à une sélection indirecte parce qu'elle favorise la substitution postérieure d'une autre mutation bénéfique (Woods et al 2011). Dans la population Ara-3, l'innovation phénotypique conduisant à la capacité à utiliser le citrate (Blount et al 2008) dépend d'une succession de mutations, c'est-à-dire d'événements de contingence historique, même si les bases génétiques ne sont pas toutes élucidées sauf pour la dernière étape mutationnelle (Blount et al 2012). D'autre part, l'évolution vers des niveaux élevés de résistance aux antibiotiques implique également des trajectoires mutationnelles spécifiques (Weinreich et al 2006, Almahmoud et al 2009).

L'émergence du polymorphisme dans la population Ara-2 est elle aussi associée à une succession de mutations. En effet, dans la lignée S, les trois mutations des gènes *spoT*, *arcA* et *gntR* sont nécessaires à l'établissement des phénotypes correspondants. Nous avons par ailleurs montré que ces trois mutations étaient substituées dans cet ordre dans la population Ara-2. Par ailleurs, cet ordre des mutations est probablement important dans l'émergence de la lignée S. En effet, l'addition successive de chacune des trois mutations dans le contexte génétique ancestral entraîne une augmentation progressive de la valeur sélective (voir Résultats Chapitre II), alors que la combinaison des deux mutations *spoT* et *gntR* sans la mutation *arcA*, conduit à une diminution de la valeur sélective (résultats non montrés). L'émergence de la lignée S est donc contingente de son histoire évolutive passée.

Cette contingence historique présente probablement un niveau de complexité plus important encore, dans la mesure où l'émergence de la lignée S est sans doute également dépendante de l'émergence de la lignée L. En effet, nos résultats montrent

que les phénotypes associés à la lignée L (interactions dépendantes de la fréquence et sécrétion de molécules dans le surnageant de culture) se mettent en place de façon graduelle au cours du temps, tout comme pour la lignée S (voir Résultats Supplémentaires Partie II). Cette double contingence historique, intra- et inter-lignée participe probablement à la rareté, mais également à la complexité, de ce polymorphisme.

IV. Perspectives

L'objectif majeur de ce travail de thèse était d'identifier les bases génétiques de l'émergence du polymorphisme. Il a été mené à bien pour la lignée S, pour laquelle les mutations *spoT*, *arcA* et *gntR* sont nécessaires et suffisantes à la totalité des phénotypes connus à ce jour. Trois axes majeurs peuvent être envisagés pour poursuivre ces travaux. Le premier consistera à comprendre les effets métaboliques, moléculaires et physiologiques de ces mutations qui affectent des gènes régulateurs. En particulier, elles peuvent entraîner des variations des taux de (p)ppGpp et des niveaux d'expression (et donc d'activités) des enzymes impliquées dans le cycle de Krebs, le shunt du glyoxylate et la voie Entner-Doudoroff. La réorientation des flux métaboliques devra ainsi être analysée. Nous réaliserons également des profils globaux de transcription pour les différentes souches isogéniques, avec et sans ces trois mutations, construites et décrites dans le chapitre II des Résultats. Ceci permettra notamment de déterminer la dynamique des changements globaux de transcription au cours de l'émergence de la lignée S, mais aussi le régulon dépendant de chacun de ces trois

régulateurs globaux. Enfin, les analyses biochimiques pour caractériser les propriétés de la protéine ArcA évoluée (liaison à l'ADN, phosphorylation, dimérisation) par rapport à celles de la protéine ancestrale seront poursuivies (voir Résultats Chapitre III). Les mêmes analyses biochimiques pourront être réalisées pour les protéines GntR ancestrale et évoluée.

Le second axe consistera à comprendre les différences phénotypiques conférées par les différents allèles des gènes *spoT* et *arcA*. Des expériences similaires à celles proposées ci-dessus (analyses métaboliques, de transcription globale et biochimiques) seront réalisées.

Le troisième axe consistera à identifier les bases génétiques conduisant à la lignée L. Les analyses des génomes en cours de séquençage devraient permettre de mettre en évidence les mutations impliquées. Des expériences similaires à celles présentées dans les Chapitres II et III des Résultats devraient alors permettre de répondre aux questions suivantes : les contraintes génétiques qui pèsent sur l'émergence de la lignée S sont-elles aussi élevées pour la lignée L ? Quelle est la nature exacte des interactions entre S et L ? Quel est le substrat relargué par les cellules L dans le milieu de culture que les cellules S peuvent utiliser pour leur croissance ? Les mutations conduisant à la lignée S sont-elles contingentes de celles conduisant à la lignée L ?

Les réponses à toutes ces questions devraient alors permettre la compréhension complète de l'émergence de ce polymorphisme et, de façon générale, de la mise en place d'interactions entre les êtres vivants, conduisant à l'étape ultime de diversification adaptative.

REFERENCES

- Almahmoud I, Kay E, Schneider D and Maurin M. 2009. Mutational paths towards increased fluoroquinolone resistance in *Legionella pneumophila*. J Antimicrob Chemother. 64:284-93.
- Amikam D and Galperin MY. 2006. PilZ domain is part of the bacterial c-di-GMP binding protein. Bioinformatics 22:3-6.
- Applebee MK, Herrgård MJ and Palsson BO. 2008. Impact of individual mutations on increased fitness in adaptively evolved strains of *Escherichia coli*. Journal of Bacteriology 190:5087–5094.
- Atlas R.M. 2004. Handbook of Microbiological Media (CRC Press, Boca Raton, FL)
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL and Mori H. 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2:2006-0008.
- Babitzke P and Romeo T. 2007. *CsrB* sRNA family: sequestration of RNA-binding regulatory proteins. Current Opinion in Microbiology 10:156-163.
- Ball CA, Osuna R, Ferguson KC, Johnson RC. 1992. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. Journal of Bacteriology 174:8043-8056.
- Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ and Rainey PB. 2007. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. Genetics 176:441-453.
- Barker MM, Gaal T, Gourse RL. 2001. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. Journal of Molecular Biology 305:689-702.

- Barrick JE and Lenski RE. 2009. Genome-wide mutational diversity in an evolving population of *Escherichia coli*. Cold Spring Harbor Symposia on Quantitative Biology 74:119-129.
- Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE and Kim JF. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. Nature 461:1243-1247.
- Battesti A, Majdalani N and Gottesman S. 2011. The RpoS-mediated general stress response in *Escherichia coli*. Annual Review of Microbiology 65:189-213.
- Beaumont HJ, Gallie J, Kost C, Ferguson GC and Rainey PB. 2009. Experimental evolution of bet hedging. Nature. 462:90-3.
- Besier S, Zander J, Kahl BC, Kraiczy P, Brade V and Wichelhaus TA. 2008. The thymidine-dependent small-colony-variant phenotype is associated with hypermutability and antibiotic resistance in clinical *Staphylococcus aureus* isolates. Antimicrob Agents Chemother. 52:2183-9.
- Bhagwat AA, Tan J, Sharma M, Kothary M, Low S, Tall BD and Bhagwat M. 2006. Functional heterogeneity of RpoS in stress tolerance of enterohemorrhagic *Escherichia coli* strains. Appl Environ Microbiol. 72:4978-86.
- Blount ZD, Borland CZ and Lenski RE. 2008. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proceedings of the National Academy of Sciences, USA 105:7899-7906.
- Blount ZD, Barrick JE, Davidson CJ and Lenski RE. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489:513-518.

- Bosch L, Nilsson L, Vijgenboom E and Verbeek H. 1990. Fis-dependent trans-activation of tRNA and rRNA operons of *Escherichia coli*. *Biochim Biophys Acta*.1050:293-301.
- Brown TD, Jones-Mortimer MC and Kornberg HL. 1977. The enzymic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*. *J Gen Microbiol*. 102:327-36.
- Browning DF, Grainger DC and Busby SJ. 2010. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Current Opinion in Microbiology* 13:773-780.
- Cezairliyan BO and Sauer RT. 2009. Control of *Pseudomonas aeruginosa* AlgW protease cleavage of MucA by peptide signals and MucB. *Mol Microbiol*. 72:368-79.
- Chatterjee I, Herrmann M, Proctor RA, Peters G and Kahl BC. 2007. Enhanced post-stationary-phase survival of a clinical thymidine-dependent small-colony variant of *Staphylococcus aureus* results from lack of a functional tricarboxylic acid cycle. *J Bacteriol*. 189:2936-40.
- Chin KH, Lee YC, Tu ZL, Chen CH, Tseng YH, Yang JM, Ryan RP, McCarthy Y, Dow JM, Wang AH and Chou SH. 2010. The cAMP receptor-like protein CLP is a novel c-di-GMP receptor linking cell-cell signaling to virulence gene expression in *Xanthomonas campestris*. *Journal of Molecular Biology* 396:646-62.
- Cho BK, Knight EM and Palsson BØ. 2006. Transcriptional regulation of the fad regulon genes of *Escherichia coli* by ArcA. *Microbiology* 152:2207-2219.
- Cohan FM. 2002. What are bacterial species? *Annu Rev Microbiol*. 2002;56:457-87.

- Cohan FM and Koeppel AF. 2008. The Origins of Ecological Diversity in Prokaryotes. *Current Biology* 18:1024-1034.
- Conrad TM , Frazier M, Joyce AR, Cho BK, Knight EM, Lewis NE, Landick R, Palsson BØ. 2010. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proceedings of the National Academy of Sciences, USA* 107:20500-20505.
- Conway T. 1992. The Entner-Doudoroff pathway: history, physiology and molecular biology. *FEMS Microbiol Rev.* 9:1-27.
- Cooper TF, Rozen DE and Lenski RE. 2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 100:1072-1077.
- Cooper TF, Remold SK, Lenski RE and Schneider D. 2008. Expression profiles reveal parallel evolution of epistatic interactions involving the CRP regulon in *Escherichia coli*. *PLoS Genetics* 4:e35 [10 pp].
- Cooper VS, Bennett AF and Lenski RE. 2001. Evolution of thermal dependence of growth rate of *Escherichia coli* populations during 20,000 generations in a constant environment. *Evolution* 55:889-896.
- Crozat E, Philippe N, Lenski RE, Geiselmann J and Schneider D. 2005. Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169:523-532.
- Crozat E, Winkworth C, Gaffé J, Hallin PF, Riley MA, Lenski RE and Schneider D. 2010. Parallel genetic and phenotypic evolution of DNA superhelicity in experimental populations of *Escherichia coli*. *Molecular Biology and Evolution* 27:2113-2128.

- Crozat E, Hindré T, Kühn L, Garin J, Lenski RE and Schneider D. 2011. Altered regulation of the OmpF porin by Fis in *Escherichia coli* during an evolution experiment and between B and K12 strains. *Journal of Bacteriology* 193:429-440.
- Damron FH, Owings JP, Okkotsu Y, Varga JJ, Schurr JR, Goldberg JB, Schurr MJ and Yu HD. 2012. Analysis of the *Pseudomonas aeruginosa* regulon controlled by the sensor kinase KinB and sigma factor RpoN. *J Bacteriol.* 194:1317-30.
- Darwin C. 1859. *On the Origin of Species by Means of Natural Selection*. Murray, London.
- Davis BD and Mingioli ES. 1950. Mutants of *Escherichia coli* requiring methionine or vitamin B12. *Journal of Bacteriology* 60:17–28.
- de Visser JAGM and Rozen DE. 2006. Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics* 172:2093–2100.
- Deitsch KW, Lukehart SA and Stringer JR. 2009. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat Rev Microbiol.* 7:493-503
- Dorman CJ et Deighan P. 2003. Regulation of gene expression by histone-like proteins in bacteria. *Current Opinion in Genetics and Development* 13:179-84.
- Dykhuizen DE and Dean AM. 1990. Enzyme activity and fitness: Evolution in solution. *Trends Ecol Evol.* 5:257-62.
- Ferea TL, Botstein D, Brown PO and Rosenzweig RF. 1999. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc Natl Acad Sci USA.* 96:9721-6.

- Ferenci T. 2005. Maintaining a healthy SPANC balance through regulatory and mutational adaptation. *Mol Microbiol.* 57:1-8.
- Ferenci T. 2008. Bacterial physiology, regulation and mutational adaptation in a chemostat environment. *Advances in Microbial Physiology* 53:169-229.
- Ferenci T, Galbiati HF, Betteridge T, Phan K and Spira B. 2011. The constancy of global regulation across a species: the concentrations of ppGpp and RpoS are strain-specific in *Escherichia coli*. *BMC Microbiol* 11:62.
- Fiegna F, Yu YT, Kadam SV and Velicer GJ. 2006. Evolution of an obligate social cheater to a superior cooperator. *Nature* 441:310-4.
- Finkel SE et Johnson RC. 1992. The Fis protein: it's not just for DNA inversion anymore. *Molecular Microbiology* 6:3257-65.
- Gao W, Chua K, Davies JK, Newton HJ, Seemann T, Harrison PF, Holmes NE, Rhee HW, Hong JI, Hartland EL, Stinear TP and Howden BP. 2010. Two novel point mutations in clinical *Staphylococcus aureus* reduce linezolid susceptibility and switch on the stringent response to promote persistent infection. *PLoS Pathog.* 6:e1000944.
- Geipel U. 2009. Pathogenic organisms in hip joint infections. *Int J Med Sci.* 6:234-40.
- Georgellis D, Kwon O and Lin EC. 2001. Quinones as the redox signal for the *arc* two-component system of bacteria. *Science* 292:2314-2316.
- Gerrish PJ and RE Lenski. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica* 102/103:127-144.
- Gonzalez-Gil G, Bringmann P and Kahmann R. 1996. Fis is a regulator of metabolism in *Escherichia coli*. *Molecular Microbiology* 22:21-9.

- Gottesman S. 2004. The small RNA regulators of *Escherichia coli*: roles and mechanisms. Annual Review of Microbiology 58:303-28.
- Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D and Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 4:e1000303.
- Hall B. 1982. Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. Journal of Bacteriology 151:269–273.
- Häussler S, Ziegler I, Löttel A, von Götz F, Rohde M, Wehmhöner D, Saravanamuthu S, Tümmler B and Steinmetz I. 2003. Highly adherent small-colony variants of *Pseudomonas aeruginosa* in cystic fibrosis lung infection. J Med Microbiol. 52:295-301.
- Helling RB, Vargas CN and Adams J. 1987. Evolution of *Escherichia coli* during growth in a constant environment. Genetics 116:349–358.
- Hengge R. 2008. The two-component network and the general stress sigma factor RpoS (sigma S) in *Escherichia coli*. Advances in Experimental Medicine and Biology 631:40-53.
- Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR and Palsson BØ. 2006. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nature Genetics 38:1406-1412.
- Hindré T, Knibbe C, Beslon G and Schneider D. 2012. New insights into bacterial adaptation through in vivo and in silico experimental evolution. Nature Review Microbiology 10:352-65.

- Hogardt M and Heesemann J. 2010. Adaptation of *Pseudomonas aeruginosa* during persistence in the cystic fibrosis lung. *Int J Med Microbiol.* 300:557-62.
- Huse HK, Kwon T, Zlosnik JE, Speert DP, Marcotte EM and Whiteley M. 2010. Parallel evolution in *Pseudomonas aeruginosa* over 39,000 generations in vivo. *MBio.* 1:e00199-10.
- Ishiguro N, Oka C and Sato G. 1978. Isolation of citrate positive variants of *Escherichia coli* from domestic pigeons, pigs, cattle, and horses. *Applied and Environmental Microbiology* 36:217–222.
- Ishiguro N, Oka C, Hanazawa Y and Sato G. 1979. Plasmids in *Escherichia coli* controlling citrate-utilizing ability. *Applied and Environmental Microbiology* 28:956–964.
- Iuchi S and Lin EC. 1988. *arcA* (dye), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proc Natl Acad Sci USA.* 85:1888-92.
- Iuchi S, Cameron DC and Lin EC. 1989. A second global regulator gene (*arcB*) mediating repression of enzymes in aerobic pathways of *Escherichia coli*. *J Bacteriol.* 171:868-73.
- Jenal U. 2004. Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? *Current Opinion in Microbiology* 7:185-91.
- Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi S-H, Couloux A, Lee S-W, Yoon SH, Cattolico L, Hur C-G, Park H-S, Ségurens B, Kim SC, Oh TK, Lenski RE,

- Studier FW, Daegelen P and Kim JF. 2009. Genome sequences of *Escherichia coli* B strains REL606 and BL21(DE3). *Journal of Molecular Biology* 394:644-652.
- Jishage M, Kvint K, Shingler V and Nyström T. 2002. Regulation of sigma factor competition by the alarmone ppGpp. *Genes and Development* 15:1260-1270.
- Kassen R and Rainey PB. 2004. The ecology and genetics of microbial diversity. *Annu Rev Microbiol.* 58:207-31.
- King T, Ishihama A, Kori A and Ferenci T. 2004. A regulatory trade-off as a source of strain variation in the species *Escherichia coli*. *J Bacteriol.* 186:5614-20.
- Khan AI, Dinh DM, Schneider D, Lenski RE and Cooper TF. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332:1193-6.
- Kornberg HL. 1966. The role and control of the glyoxylate cycle in *Escherichia coli*. *Biochem J.* 99:1-11.
- Koser SA. 1923. Utilization of the salts of organic acids by colon-aerogenes group. *Journal of Bacteriology* 8:493–520.
- Koser SA. 1924. Correlation of citrate-utilization by members of the colon-aerogenes group with other differential characteristic and with habitat. *Journal of Bacteriology* 9:59–77.
- Laland KN, Odling-Smee FJ and Feldman MW. 1999. Evolutionary consequences of niche construction and their implications for ecology. *Proc Natl Acad Sci U S A.* 96:10242-7.
- Lannergard J, von Eiff C, Sander G, Cordes T, Seggewiss J, Peters G, Proctor RA, Becker K and Hughes D. 2008. Identification of the genetic basis for clinical

menadione-auxotrophic small-colony variant isolates of *Staphylococcus aureus*.

Antimicrob Agents Chemother.52:4017-22.

Lara F and Stokes J. 1952. Oxidation of citrate by *Escherichia coli*. Journal of Bacteriology 9:59–77.

Lazarus LR and Travers AA. 1993. The *Escherichia coli* FIS protein is not required for the activation of *tyrT* transcription on entry into exponential growth. EMBO Journal 12:2483-94.

Lee DH and Palsson BØ. 2010. Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. Appl Environ Microbiol. 76:4158-68.

Le Gac M, Brazas MD, Bertrand M, Tyerman JG, Spencer CC, Hancock RE and Doebeli M. 2008. Metabolic changes associated with adaptive diversification in *Escherichia coli*. Genetics. 178:1049-60.

Le Gac M and Doebeli M. 2010. Epistasis and frequency dependence influence the fitness of an adaptive mutation in a diversifying lineage. Molecular Ecology. 19:2430-8.

Le Gac M, Plucain J, Hindré T, Lenski RE and Schneider D. 2012. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. Proceedings of the National Academy of Sciences, USA 109:9487-9492.

Lenski RE, Rose MR, Simpson SC and Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. American Naturalist 138:1315-1341.

- Lenski RE, and Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences, USA* 91:6808-6814.
- Lenski RE, Winkworth CL and Riley MA. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *Journal of Molecular Evolution* 56:498-508.
- Lenski RE. 2004. Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*. *Plant Breeding Reviews*
- Lenski RE, Barrick JE and Ofria C. 2006. Balancing robustness and evolvability (Essay). *PLoS Biology* 4:e428 [3 pp].
- Leveret M, Zamfir O, Clermont O, Bouvet O, Lespinats S, Hipeaux MC, Branger C, Picard B, Saint-Ruf C, Norel F, Balliau T, Zivy M, Le Nagard H, Cruveiller S, Chane-Woon-Ming B, Nilsson S, Gudelj I, Phan K, Ferenci T, Tenaillon O and Denamur E. 2010. Molecular and evolutionary bases of within-patient genotypic and phenotypic diversity in *Escherichia coli* extraintestinal infections. *PLoS Pathog.* 6:e1001125.
- Lieberman TD, Michel JB, Aingaran M, Potter-Bynoe G, Roux D, Davis MR Jr, Skurnik D, Leiby N, LiPuma JJ, Goldberg JB, McAdam AJ, Priebe GP and Kishony R. 2011. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet.* 43:1275-80.
- Luijsterburg MS, Noom MC, Wuite GJ and Dame RT. 2006. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *Journal of Structural Biology* 156:262-72.

- Lynch AS and Lin ECC. 1996. Response to molecular oxygen. *Escherichia coli* and *Salmonella*, cellular and molecular biology. ASM press, Washington D.C. vol1:1526-1538
- Magnusson LU, Farewell A and Nyström T. 2005. ppGpp: a global regulator in *Escherichia coli*. Trends in Microbiology 13:236-42.
- Maharjan R, Zhou Z, Ren Y, Li Y, Gaffé J, Schneider D, McKenzie C, Reeves PR, Ferenci T and Wang L. 2010. Genomic identification of a novel mutation in *hfq* that provides multiple benefits in evolving glucose-limited populations of *Escherichia coli*. Journal of Bacteriology 192:4517-4521.
- Malone JG, Williams R, Christen M, Jenal U, Spiers AJ and Rainey PB. 2007. The structure-function relationship of WspR, a *Pseudomonas fluorescens* response regulator with a GGDEF output domain. Microbiology. 153:980-94.
- Malone JG, Jaeger T, Spangler C, Ritz D, Spang A, Arrieumerlou C, Kaeffer V, Landmann R and Jenal U. 2010. YfiBNR mediates cyclic di-GMP dependent small colony variant formation and persistence in *Pseudomonas aeruginosa*. PLoS Pathog. 2010 Mar 12;6(3):e1000804.
- Malpica R, Franco B, Rodriguez C, Kwon O and Georgellis D. 2004. Identification of a quinone-sensitive redox switch in the ArcB sensor kinase. Proc Natl Acad Sci U S A. 101:13318-23.
- Malpica R, Sandoval GR, Rodríguez C, Franco B and Georgellis D. 2006. Signaling by the *arc* two-component system provides a link between the redox state of the quinone pool and gene expression. Antioxidants and Redox Signaling 8:781-95.
- Mayr, E. 1963. Animal Species and Evolution. Belknap, Cambridge, MA.

- McDonald MJ, Gehrig SM, Meintjes PL, Zhang X-X and Rainey PB. 2009. Adaptive Divergence in Experimental Populations of *Pseudomonas fluorescens*. IV. Genetic Constraints Guide Evolutionary Trajectories in a Parallel Adaptive Radiation. *Genetics* 183:1041-1053.
- Melter O and Radojevič B. 2010. Small colony variants of *Staphylococcus aureus*--review. *Folia Microbiol (Praha)*. 55:548-58.
- Mergeay M, Gigot D, Beckmann J, Glansdorff N and Piérard A. 1974. Physiology and genetics of carbamoylphosphate synthesis in *Escherichia coli* K12. *Mol Gen Genet*. 133:299-316.
- Mika F and Hengge R. 2005. A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of sigmaS (RpoS) in *E. coli*. *Genes Dev*. 19:2770-81.
- Mikkelsen H, Sivaneson M and Filloux A. 2011. Key two-component regulatory systems that control biofilm formation in *Pseudomonas aeruginosa*. *Environnemental Microbiology* 13:1666-1681.
- Newell PD, Monds RD and O'Toole GA. 2009. LapD is a bis-(3',5')-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1. *Proceedings of the National Academy of Sciences, USA* 106:3461-6.
- Ninnemann O, Koch C and Kahmann R. 1992. The *E. coli* *fis* promoter is subject to stringent control and autoregulation. *EMBO Journal* 11:1075-83.
- Notley-McRobb L, King T and Ferenci T. 2002. *rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses. *Journal of Bacteriology* 184:806-11.

- Odling-Smee FJ, Laland KN and Feldman MW. 1996. Niche construction. *Am Nat* 147:641-648.
- Ostrowski EA, Woods RJ and Lenski RE. 2008. The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. *Proc Biol Sci.* 275:277-84.
- Owens RM, Pritchard G, Skipp P, Hodey M, Connell SR, Nierhaus KH and O'Connor CD. 2004. A dedicated translation factor controls the synthesis of the global regulator Fis. *EMBO Journal* 23: 3375-3385.
- Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE and Blot M. 1999. Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences, USA* 96:3807-3812.
- Peekhaus N and Conway T. 1998. What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J Bacteriol.* 180:3495-502.
- Pelosi L, Kühn L, Guetta D, Garin J, Geiselmann J, Lenski RE and Schneider D. 2006. Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173:1851-1869.
- Pena-Sandoval GR, Kwon O and Georgellis D. 2005. Requirement of the receiver and phosphotransfer domains of ArcB for efficient dephosphorylation of phosphorylated ArcA in vivo. *Journal of Bacteriology* 187:3267-3272.
- Philippe N, Crozat E, Lenski RE and Schneider D. 2007. Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*. *BioEssays* 29:846-860.

- Philippe N, Pelosi L, Lenski RE and Schneider D. 2009. Evolution of penicillin-binding protein 2 concentration and cell shape during a long-term experiment with *Escherichia coli*. *Journal of Bacteriology* 191:909-921.
- Proctor RA, van Langevelde P, Kristjansson M, Maslow JN and Arbeit RD. 1995. Persistent and relapsing infections associated with small-colony variants of *Staphylococcus aureus*. *Clin Infect Dis*. 20:95-102.
- Proctor RA, Kahl B, von Eiff C, Vaudaux PE, Lew DP and Peters G. 1998. Staphylococcal small colony variants have novel mechanisms for antibiotic resistance. *Clin Infect Dis*. 27Suppl 1:S68-74.
- Rainey PB and Travisano M. 1998. Adaptive radiation in a heterogeneous environment. *Nature* 394:69-72.
- Rainey PB, Buckling A, Kassen R and Travisano M. 2000. The emergence and maintenance of diversity: insights from experimental bacterial populations. *Trends Ecol Evol*. 15:243-247.
- Rainey PB and Rainey K. 2003. Evolution of cooperation and conflict in experimental bacterial populations. *Nature* 425:72-4.
- Robbe-Saul V, Algorta G, Rouilhac I and Norel F. 2003. Characterization of the RpoS status of clinical isolates of *Salmonella enterica*. *Appl Environ Microbiol*. 69:4352-8.
- Robichon C, Luo J, Causey TB, Benner JS and Samuelson JC. 2011. Engineering *Escherichia coli* BL21(DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Appl Environ Microbiol*. 77:4634-46.

- Rosenzweig RF, Sharp RR, Treves DS and Adams J. 1994. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* 137:903-17.
- Ross W, Thompson JF, Newlands JT and Gourse RL. 1990. *Escherichia coli* Fis protein activates ribosomal RNA transcription in vitro and in vivo. *EMBO Journal* 9:3733-3742.
- Rozen DE and Lenski RE. 2000. Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *American Naturalist* 155:24-35.
- Rozen DE, Schneider D and Lenski RE. 2005. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. *Journal of Molecular Evolution* 61:171-180.
- Rozen DE, Philippe N, de Visser JA, Lenski RE and Schneider D. 2009. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology Letters* 12:34-44.
- Ryan RP, Tolker-Nielsen T and Dow JM. 2012. When the PilZ don't work: effectors for cyclic di-GMP action in bacteria. *Trends in Microbiology* 20:235-42.
- Ryan VT, Grimwade JE, Camara JE, Crooke E, Leonard AC. 2004. *Escherichia coli* prereplication complex assembly is regulated by dynamic interplay among Fis, IHF and DnaA. *Molecular Microbiology* 51:1347-1359.
- Salmon KA, Hung SP, Steffen NR, Krupp R, Baldi P, Hatfield GW and Gunsalus RP. 2005. Global gene expression profiling in *Escherichia coli* K12: effects of oxygen availability and ArcA. *Journal of Biological Chemistry* 280:15084-15096.

- Sambrook J, Fritsch EF and Maniatis T. 1989. Molecular cloning: A laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sappington KJ, Dandekar AA, Oinuma K and Greenberg EP. 2011. Reversible signal binding by the *Pseudomonas aeruginosa* quorum-sensing signal receptor LasR. MBio. 2:e00011-11.
- Scheutz F and Strockbine NA. 2005. Genus I. *Escherichia*, *Castellani* and *Chalmers* 1919. In (Ed.) Garrity, G.M., Brenner, D., Kreig, N.R., Staley, J.R. Bergey's Manual of Systematic Bacteriology, Volume Two: The Proteobacteria (Springer, New York, NY).
- Schneider D, Duperchy E, Coursange E, Lenski RE and Blot M. 2000. Long-term experimental evolution in *Escherichia coli*. IX. Characterization of IS-mediated mutations and rearrangements. Genetics 156:477-488.
- Schneider R, Travers A, Muskhelishvili G. 2000. The expression of the *Escherichia coli* *fis* gene is strongly dependent on the superhelical density of DNA. Molecular Microbiology 38:167-175.
- Seeto S, Notley-McRobb L and Ferenci T. 2004. The multifactorial influences of RpoS, Mlc and cAMP on *ptsG* expression under glucose-limited and anaerobic conditions. Res Microbiol. 155:211-5.
- Shin D, Cho N, Heu S and Ryu S. 2003. Selective regulation of *ptsG* expression by Fis. Formation of either activating or repressing nucleoprotein complex in response to glucose. Journal of Biological Chemistry 278:14776-14781.

- Shin JS, Ryu KS, Ko J, Lee A and Choi BS. 2011. Structural characterization reveals that a PilZ domain protein undergoes substantial conformational change upon binding to cyclic dimeric guanosine monophosphate. *Protein Sci.* 20:270-7.
- Sniegowski PD, Gerrish PJ and Lenski RE. 1997. Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* 387:703-705.
- Spencer CC, Bertrand M, Travisano M and Doebeli M. 2007. Adaptive diversification in genes that regulate resource use in *Escherichia coli*. *PLoS Genet.* 3:e15.
- Spiers AJ, Kahn SG, Bohannon J, Travisano M and Rainey PB. 2002. Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics*. 161:33-46.
- Spiers AJ, Bohannon J, Gehrig SM and Rainey PB. 2003. Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Mol Microbiol.* 50:15-27.
- Spira B, Hu X and Ferenci T. 2008. Strain variation in ppGpp concentration and RpoS levels in laboratory strains of *Escherichia coli* K-12. *Microbiology* 154:2887-2895.
- Stanek MT, Cooper TF and Lenski RE. 2009. Identification and dynamics of a beneficial mutation in a long-term evolution experiment with *Escherichia coli*. *BMC Evolutionary Biology* 9:302
- Storz G, Vogel J and Wassarman KM. 2011. Regulation by small RNAs in bacteria: expanding frontiers. *Molecular Cell* 43:880-891.
- Studier FW, Daegelen P, Lenski RE, Maslov S and Kim JF. 2009. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and

- BL21(DE3) and comparison of the *E. coli* B and K-12 genomes. *Journal of Molecular Biology* 394:653-680.
- Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH and Breaker RR. 2008. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science*. 2008 Jul 18;321(5887):411-3.
- Tamayo R, Pratt JT and Camilli A. 2007. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. *Annual Review of Microbiology* 61:131-48.
- Tenaillon O, Skurnik D, Picard B and Denamur E. 2010. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol*. 8:207-17.
- Tonegawa S. 1983. Somatic generation of antibody diversity. *Nature* 302:575-581.
- Toro-Roman A, Mack TR and Stock AM. 2005. Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face. *J Mol Biol*. 349:11-26.
- Travers A, Schneider R and Muskhelishvili G. 2001. DNA supercoiling and transcription in *Escherichia coli*: The FIS connection. *Biochimie* 83:213-7.
- Treves DS, Manning S and Adams J. 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol Biol Evol*. 15:789-97.
- Turner PE, Souza V and Lenski RE. 1996. Tests of ecological mechanism promoting the stable coexistence of two bacterial genotypes. *Ecology* 77:2119-2129.
- Velicer GJ, Yu YT. 2003. Evolution of novel cooperative swarming in the bacterium *Myxococcus xanthus*. *Nature*. 425:75-8.

- Waegeman H, Beauprez J, Moens H, Maertens J, De Mey M, Foulquié-Moreno MR, Heijnen JJ, Charlier D and Soetaert W. 2011. Effect of *iclR* and *arcA* knockouts on biomass formation and metabolic fluxes in *Escherichia coli* K12 and its implications on understanding the metabolism of *Escherichia coli* BL21 (DE3). BMC Microbiol. 11:70.
- Wang L, Spira B, Zhou Z, Feng L, Maharjan RP, Li X, Li F, McKenzie C, Reeves PR and Ferenci T. 2010. Divergence involving global regulatory gene mutations in an *Escherichia coli* population evolving under phosphate limitation. Genome Biology and Evolution 2:478-487.
- Weinreich DM, Delaney NF, Depristo MA and Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. Science. 312:111-4.
- Weinstein-Fischer D, Elgrably-Weiss M and Altuvia S. 2000. *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. Molecular Microbiology 35:1413-20.
- Wichman HA, Badgett MR, Scott LA, Boulianne CM and Bull JJ. 1999. Different trajectories of parallel evolution during viral adaptation. Science. 285:422-4.
- Wolfe AJ. 2005. The acetate switch. Microbiol Mol Biol Rev. 69:12-50.
- Woods R, Schneider D, Winkworth CL, Riley MA and Lenski RE. 2006. Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. Proceedings of the National Academy of Sciences, USA 103:9107-9112.
- Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR and Lenski RE. 2011. Second-order selection for evolvability in a large *Escherichia coli* population. Science 331:1433-1436.

- Yang L, Jelsbak L, Marvig RL, Damkiær S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, Høiby N, Sommer MO and Molin S. 2011. Evolutionary dynamics of bacteria in a human host environment. *Proc Natl Acad Sci U S A*. 108:7481-6.
- Yanisch-Perron C, Vieira J and Messing J. 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*. 33:103-19.
- Yu YT, Yuan X and Velicer GJ. 2010. Adaptive evolution of an sRNA that controls *Myxococcus* development. *Science* 328:993.
- Zambrano MM, Siegele DA, Almirón M, Tormo A and Kolter R. 1993. Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. *Science*. 259:1757-60.
- Zinser ER and Kolter R. 2000. Prolonged stationary-phase incubation selects for *lrp* mutations in *Escherichia coli* K-12. *J Bacteriol*. 182:4361-5.
- Zhou YN and Jin DJ. 1998. The *rpoB* mutants destabilizing initiation complexes at stringently controlled promoters behave like "stringent" RNA polymerases in *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* 95:2908-13.

BASES GENETIQUES ET ECOLOGIQUES DE LA DIVERSIFICATION ADAPTATIVE CHEZ *ESCHERICHIA COLI*

Résumé : Les processus de diversification adaptative, qui sont au cœur de la diversité du monde vivant, ont été étudiés grâce à une stratégie d'évolution expérimentale, initiée par le Pr Richard Lenski en 1988. Douze populations, fondées à partir d'un ancêtre commun d'*Escherichia coli*, sont propagées indépendamment depuis plus de 55 000 générations par transferts journaliers dans un milieu minimum limité en glucose. Un événement de diversification a émergé après 6500 générations d'évolution dans une seule des douze populations, appelée Ara-2, conduisant à deux lignées cellulaires différenciées, appelées S et L, qui continuent de co-exister depuis notamment grâce à des interactions négatives dépendant de leur fréquence. Deux propriétés confèrent à ce polymorphisme une grande originalité et donc un intérêt d'étude important : sa durée car il s'agit du plus long polymorphisme jamais identifié lors d'expériences d'évolution en laboratoire, et son unicité puisqu'il ne s'est produit qu'une seule fois au sein des douze populations initiées à partir d'un ancêtre commun.

L'objectif de ce travail a été d'identifier les mécanismes du maintien au long terme des lignées S et L, ainsi que les bases génétiques de leur émergence. Le maintien du polymorphisme est lié à une forte dynamique des relations écologiques entre S et L, l'une des lignées envahissant systématiquement les niches écologiques de l'autre, qui réagit en conséquence pour éviter l'extinction. L'émergence de la lignée S est due à une succession précise de trois mutations, nécessaires et suffisantes pour établir les phénotypes de la lignée S. Les trois mutations affectent toutes des gènes codant des régulateurs globaux de la transcription, dont deux sont impliqués dans la régulation du métabolisme central. Pour l'un d'entre eux, l'allèle évolué altère les propriétés de liaison à l'ADN de la protéine évoluée. Bien que ce polymorphisme soit unique, ces trois gènes sont pourtant les cibles de la sélection naturelle dans la majorité des autres populations de l'expérience d'évolution. Pour deux d'entre eux, seul l'allèle substitué dans la population Ara-2 confère en fait les phénotypes de la lignée S. Ainsi, l'unicité de cet événement de diversification est liée à une succession d'événements mutationnels très précis, qui affectent par ailleurs les réseaux globaux de l'expression des gènes. Ces modifications graduelles ont ainsi conduit à l'émergence du plus long polymorphisme mis en évidence à ce jour dans des expériences d'évolution en laboratoire.